



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/82, C07K 14/415, C12Q 1/68, C12N 1/20, A01H 5/00, C07K 16/16		A1	(11) International Publication Number: WO 00/44918 (43) International Publication Date: 3 August 2000 (03.08.00)
(21) International Application Number: PCT/GB00/00248 (22) International Filing Date: 28 January 2000 (28.01.00)		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(30) Priority Data: 9901927.5 28 January 1999 (28.01.99) GB		Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(71) Applicant (for all designated States except US): PLANT BIOSCIENCE LIMITED [GB/GB]; Norwich Research Park, Colney Lane, Norwich, Norfolk NR4 7UH (GB). (72) Inventors; and (75) Inventors/Applicants (for US only): DEAN, Caroline [GB/GB]; 19 Waverley Road, Norwich, Norfolk NR4 6SG (GB). GENDALL, Anthony [GB/GB]; 167 Unthank Road, Norwich, Norfolk NR2 2PG (GB). (74) Agents: WALTON, Seán, M. et al.; Mewburn Ellis, York House, 23 Kingsway, London WC2B 6HP (GB).			
(54) Title: METHODS AND MEANS FOR MODIFICATION OF PLANT CHARACTERISTICS USING THE VERNALIZATION GENE VRN2			
(57) Abstract An isolated nucleic acid obtainable from the VRN2 locus of a plant, which nucleic acid encodes a polypeptide which is capable or affecting one or more physical characteristics of a plant into which the nucleic acid is introduced, the physical characteristics being selected from vernalization response, flowering time, leaf size and/or shape or shade avoidance response; alleles, fragment and derivatives thereof; polypeptides encoded by such nucleic acids; antibodies to such polypeptides.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	MW	Malawi	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Mexico	US	United States of America
CA	Canada	IT	Italy	MX	Nicaragua	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakhstan	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

METHODS AND MEANS FOR MODIFICATION OF PLANT CHARACTERISTICS USING THE VERNALIZATION GENE VRN2

The present invention relates generally to modification of characteristics of plants, especially vernalization response, flowering time, leaf shape and/or shade avoidance response, and has arisen on the basis of cloning of the VRN2 gene and mutant alleles of *Arabidopsis thaliana*, and identification of homologues in other species.

10

In different embodiments, the present invention provides for manipulation of flowering time and/or other characteristics of plants, e.g. by up or down regulating VRN2 gene expression.

The present invention also provides for modification of the extent of alteration of a relevant plant characteristic through the use of gene alleles, mutants and variants.

Plants must integrate a wide variety of environmental signals in order to maximize their reproductive success. Part of this integration must involve perception of the seasons, both to ensure the plant flowers during the correct season (for which it is adapted) and to synchronise its flowering with other members of its own species, to increase the chances of cross-fertilization. *Arabidopsis thaliana* serves as a model plant, for it exhibits responses to a wide variety of environmental stimuli that are observed in many species. Flowering in naturally occurring strains (ecotypes) of *Arabidopsis* can be promoted either by long days (increased photoperiod) or by vernalization, a long cold treatment that mimics the cold of winter. While many aspects of the photoperiodic response are now understood, the vernalization pathway has received relatively less attention. The inventors have used a late flowering, vernalization responsive mutant of *Arabidopsis*, the *fca* mutant, as a background in which to isolate mutants that exhibit a reduced vernalization response, the VRN mutants.

Vernalization is the low temperature promotion of flowering. It can also be thought of as the cold aspect of the winter season, which also includes reduced daylight hours. Many species of plants that grow in temperate or cooler climes have 5 an obligate requirement for vernalization in order to flower. Such plants typically germinate in autumn, and over-winter as vegetative plants, and flower in milder conditions of spring. Vernalization thus acts as a cue, to allow plants to sense the seasons, and to time their flowering to maximise their chance 10 of reproductive success.

Species for which flowering is important to crop production are numerous, essentially all crops which are grown from seed, with important examples being the cereals, rice and maize, 15 probably the most agronomically important in warmer climatic zones, and wheat, barley, oats and rye in more temperate climates. Important seed products are oil seed rape, sugar beet, maize, sunflower, soybean and sorghum. Many crops which are harvested for their roots are, of course, grown annually 20 from seed and the production of seed of any kind is very dependent upon the ability of the plant to flower, to be pollinated and to set seed. In horticulture, control of the timing of flowering is important. Horticultural plants whose flowering may be controlled include lettuce, endive and 25 vegetable brassicas including cabbage, broccoli and cauliflower, and carnations and geraniums.

Arabidopsis thaliana is a facultative long day plant, flowering early under long days and late under short days. 30 Because it has a small, well-characterized genome, is relatively easily transformed and regenerated and has a rapid growing cycle, *Arabidopsis* is an ideal model plant in which to study flowering and its control.

35 In addition to cloning the VRN2 gene, the inventors have unexpectedly found indication that VRN2 is a transcription factor, which in itself opens several exciting avenues for application of the present invention. Without being bound by

theory or in anyway limiting the scope of the present invention, VRN2 may be required for a normal vernalization response because it acts as a regulator of genes that ultimately lead to the transition from vegetative to reproductive growth. In such a model, cold, or a downstream molecule involved in cold perception, may regulate the activity or expression of the VRN2 protein, which in turn may regulate the expression of a large number of genes that ultimately lead to flowering. Furthermore, the shade avoidance phenotype exhibited by the *vrn2-1* mutant, as demonstrated experimentally below, provides indication that VRN2 also plays a role in regulating leaf shape, particularly in response to increased far-red light. Together, these two processes affected by a deficiency or reduction of VRN2 activity provide for a number of approaches of agronomic interest.

First, forced expression of VRN2 (for example under the control of a strong and constitutive promoter, such as the Ca MV 35 S promoter) in a wild-type background may be used alter the vernalization requirement of a plant prior to flowering. As a large number of commercial cultivars of several species, including (diploid) wheat, barley, and sugar beet, have a requirement for vernalization to flower, modification of this requirement, by reducing the duration of vernalization required, or changing the optimum temperature, or abrogating the requirement altogether, is of agronomic usefulness. For instance, a winter crop that can be sown and left in the ground for a shorter period than usual (i.e. a reduced vernalization time) may benefit from reduced risk of damage associated with severe winter weather conditions, as the crops are exposed to winter conditions for a shorter time.

Second, down-regulation of VRN2 expression, for instance by means of an antisense VRN2 cDNA, may be used to recapitulate the reduction in shade avoidance phenotype observed in *vrn2-1* mutants. This may be used in situations where crowding of the crop is a problem. Based on experimental evidence provided

herein on the phenotype of *vrn2-1* mutants, such plants are expected to exhibit less of a response to such conditions, and to produce leaves that are essentially normal i.e. as if they had not been grown in dense or crowded conditions. The normal shade avoidance phenotype is a reduction in leaf size, which reduces shade in overcrowded conditions; *vrn2-1* mutants, defective in VRN2 production, show less reduction in leaf size under conditions which would normally lead to the shade avoidance phenotype. This effect can therefore be reproduced for example by using antisense VRN2 cDNA to downregulate VRN2 expression, preventing or reducing the leaf avoidance response even in overcrowded conditions.

Third, the individual isolated domains of the VRN2 protein may be used in their own right. DNA binding of the zinc finger of VRN2 may be used to direct or control gene expression in a precise manner. The VRN2 zinc finger may recognize specific DNA sequences that represent elements in the promoters of its normal target genes. By creating fusion proteins, comprising the DNA binding (zinc finger) domain of VRN2, and an activation or repression domain from a heterologous protein, the expression of VRN2 target genes may be controlled. This allows for a precise control of the expression of those genes that are normally targets of VRN2. Given that such genes, alone or in combination, ultimately control the transition to flowering (usually following vernalization), their directed expression in other conditions may also be used to elicit changes in flowering and/or one or more other plant characteristics. The expression of (normally far-red responsive) target genes also be controlled using VRN2 fusion proteins containing the zinc finger of VRN2. Furthermore, the use of the zinc finger domain of VRN2 in conventional SELEX or one-hybrid experiments may be used to reveal the target genes or DNA sequences normally bound by VRN2.

35

The acidic activation domain of VRN2 may be used to regulate the activity of a fusion protein, including a DNA-binding protein of known specificity, and the activation domain of

VRN2. This allows for regulation of target genes of other DNA binding proteins involved in flowering, or of target genes in completely unrelated processes.

- 5 The inventors have cloned, characterised and manipulated the VRN2 gene of *Arabidopsis thaliana*, both *Columbia* and *Landsberg erecta* types, and identified alternatively spliced and mutant forms, also homologues in other species.
- 10 In the light of the inventors' experimental work, a first aspect of the present invention provides a nucleic acid isolate encoding a polypeptide including a VRN2 amino acid sequence shown herein (e.g. SEQ ID NO: 2; SEQ ID NO: 5), which may include a coding sequence shown herein (e.g. SEQ ID NO: 1; 15 SEQ ID NO: 4).

Allelic forms and alternatively spliced forms of the gene have been identified. Such polypeptides and encoding nucleic acid (e.g. as in SEQ ID NO: 8, encoded by SEQ ID NO: 7) are each 20 further provided as an aspect of the invention, as are polypeptides and nucleic acid including the mutations identified herein.

Nucleic acid according to the present invention may have the 25 sequence of the VRN2 gene of *Arabidopsis thaliana* as indicated in SEQ ID NO: 1, SEQ ID NO: 3 (*Landsberg erecta* genomic sequence), SEQ ID NO: 4 or SEQ ID NO: 6 (*Columbia* genomic sequence), or be a mutant, variant, derivative or allele or a homologue of the sequence provided. Preferred mutants, 30 variants, derivatives and alleles are those which encode a protein which retains a functional characteristic of the protein encoded by the wild-type gene, especially the ability to alter vernalization response, flowering time, leaf shape and/or shade avoidance response.

35

A mutant, variant, derivative, allele or homologue in accordance with the present invention may have the ability to

affect a physical characteristic of a plant, especially vernalization response, flowering time, leaf shape and/or shade avoidance response, as discussed.

- 5 Polynucleotides which are not 100% identical to the sequences shown herein but fall within the scope of the invention can be obtained in a number of ways.

Other VRN2 variants (for example allelic forms) of the gene
10 described herein may be obtained for example by probing cDNA or genomic DNA libraries made from *Arabidopsis thaliana* plants or cells.

In addition, other plant, monocot or dicot, homologues of the
15 gene may be obtained. Such sequences may be obtained by making or obtaining cDNA libraries made from dividing cells or tissues or genomic DNA libraries from other plant species, and probing such libraries with probes comprising all or part of a nucleic acid of the invention under conditions of medium to
20 high stringency (for example for hybridization on a solid support (filter) overnight incubation at 42°C in a solution containing 50% formamide, 5xSSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulphate and 20 µg/ml salmon sperm DNA,
25 followed by washing in 0.03M sodium chloride and 0.03M sodium citrate (i.e. 0.2x SSC) at from about 50°C to about 60°C).

Thus the present invention provides an isolated nucleic acid which hybridizes to the nucleotide sequence shown in a figure
30 herein under the abovementioned hybridization and washing conditions. Such a nucleic acid is suitable for use as a probe for detecting the VRN2 gene, for example in Southern blots.

- 35 Suitable probe and primer sequences are disclosed herein.

Alternatively, polynucleotides of the invention may be obtained by site directed mutagenesis of the sequences of

shown in the figures or allelic variants thereof. This may be useful where for example silent codon changes are required to sequences to optimise codon preferences for a particular host cell in which the polynucleotide sequences are being
5 expressed. Other sequence changes may be desired in order to introduce restriction enzyme recognition sites, or to alter the property or function of the polypeptides encoded by the polynucleotides. Further changes may be desirable to represent particular coding changes which are required to
10 provide, for example, conservative substitutions.

In the context of cloning, it may be necessary for one or more gene fragments to be ligated to generate a full-length coding sequence. Also, where a full-length encoding nucleic acid
15 molecule has not been obtained, a smaller molecule representing part of the full molecule, may be used to obtain full-length clones. Inserts may be prepared from partial cDNA clones and used to screen cDNA libraries. The full-length clones isolated may be subcloned into expression vectors and
20 activity assayed by transfection into suitable host cells, e.g. with a reporter plasmid.

The present invention also extends to nucleic acid comprising transcriptional control sequences for the VRN2 gene. Such
25 control sequences will be found 5' to the open reading frame of the gene and are obtainable by probing a genomic DNA library with a nucleic acid of the invention, selecting a clone which hybridizes under conditions of medium to high stringency, and sequencing the clone 5' to the open reading
30 frame of the gene. Where only a small amount of sequence is present in the 5' region, this sequence may be used to reprobe the library to genome walk further upstream. Analysis of the upstream region will reveal control regions for gene expression including control regions common to many genes (i.e.
35 TATA and CAAT boxes) and other control regions, usually located from 1 to 10,000, such as 1 to 1000 or 50 to 500 nucleotides upstream of the start of transcription.

To confirm that such regions are the control regions of the gene, they may be linked to a reported gene (such as β -galactosidase) and tested in any suitable *in vitro* or *in vivo* system. For example the construct of the control region (e.g. comprising 50 to 500 nucleotides upstream of the start of transcription) and the reporter gene may be used to produce a transgenic plant and the pattern of expression, both spatially and developmentally, may be compared with that of the VRN2 gene. Where substantially similar patterns of expression are found, this shows that the construct comprises substantially all of the control region of the wild type gene.

SEQ ID NO: 3 and SEQ ID NO: 6 show the nucleotide sequence of the VRN2 genomic region including promoter, respectively for *Landsberg erecta* and *Columbia* ecotypes of *Arabidopsis thaliana*, also 3' regulatory elements.

The control region may be mutated to identify specific subregions responsible for transcriptional control. This may be achieved by a number of techniques well known in the art as such, including DNase protection footprint assays, in which the control region is brought into contact with an extract from a cell in which the VRN2 gene is actively expressed, and the regions of the control region which bind factors in that extract is determined.

Isolated nucleic acid comprising such control regions obtainable by such a method form a further aspect of the present invention.

The present invention further extends to genomic DNA exon sequences found between the introns encoding a VRN2 gene in plant. Such exon sequences may be obtained in a manner analogous to that described above for the transcriptional control sequences, with the appropriate genome walking being conducted between the intron sequences. The locations of the exons may be determined by comparing genomic and cDNA

sequences of the gene, observing where the sequences line up and diverge, and looking for consensus splice sequences which define intron/exon boundaries.

- 5 As noted above, changes to a sequence, to produce a mutant, variant or derivative, may be by one or more of addition, insertion, deletion or substitution of one or more nucleotides in the nucleic acid, leading to the addition, insertion, deletion or substitution of one or more amino acids in the
10 encoded polypeptide. Of course, changes to the nucleic acid which make no difference to the encoded amino acid sequence ("degeneratively equivalent") are included.

Preferred nucleic acid sequences according to the present
15 invention are shown herein, for instance see SEQ ID NO: 1 and SEQ ID NO: 4, of which the respective predicted encoded amino acid sequences of polypeptides according to the present invention are shown in SEQ ID NO: 2 and SEQ ID NO: 5.

- 20 A mutant, allele, variant or derivative amino acid sequence in accordance with the present invention may include within a sequence shown herein a single amino acid change with respect to the sequence shown with the relevant SEQ ID NO: or in the relevant figure, or 2, 3, 4, 5, 6, 7, 8, or 9 changes, about
25 10, 15, 20, 30, 40 or 50 changes, or greater than about 50, 60, 70, 80 or 90 changes. In addition to one or more changes within the amino acid sequence shown in the relevant figure, a mutant, allele, variant or derivative amino acid sequence may include additional amino acids at the C-terminus and/or N-
30 terminus.

A sequence related to a sequence specifically disclosed herein shares homology with that sequence. Homology may be at the nucleotide sequence and/or amino acid sequence level.

- 35 Preferably, the nucleic acid and/or amino acid sequence shares homology with the coding sequence or the sequence encoded by a nucleotide sequence shown herein, for instance SEQ ID NO: 2 or SEQ ID NO: 5, preferably at least about 50%, or 60%, or 70%,

or 80% homology, most preferably at least about 90%, 95%, 96%, 97%, 98% or 99% homology.

As is well-understood, homology at the amino acid level is generally in terms of amino acid similarity or identity. Similarity allows for "conservative variation", i.e. substitution of one hydrophobic residue such as isoleucine, valine, leucine or methionine for another, or the substitution of one polar residue for another, such as arginine for lysine, glutamic for aspartic acid, or glutamine for asparagine. Similarity may be as defined and determined by the TBLASTN program, of Altschul et al. (1990) *J. Mol. Biol.* 215: 403-10, which is in standard use in the art, or, and this may be preferred, either of the standard programs BestFit and GAP, which are part of the Wisconsin Package, Version 8, September 1994, (Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA, Wisconsin 53711). BestFit makes an optimal alignment of the best segment of similarity between two sequences. Optimal alignments are found by inserting gaps to maximize the number of matches using the local homology algorithm of Smith and Waterman (Advances in Applied Mathematics (1981) 2, pp. 482-489). GAP uses the Needleman and Wunsch algorithm to align two complete sequences that maximizes the number of matches and minimizes the number of gaps. Generally, the default parameters are used, with a gap creation penalty = 12 and gap extension penalty = 4.

Homology is generally over the full-length of the relevant sequence shown herein, that is unless stated otherwise, or it may be over a contiguous sequence of about or greater than about 20, 25, 30, 33, 40, 50, 67, 133, 167, 200, 233, 267, 300, 333, 400, 450, 500, 550, 600 or more amino acids or codons, compared with the relevant amino acid sequence or nucleotide sequence as the case may be.

35

In highly preferred embodiments, all percentage homologies referred to herein refer to percentage sequence identity.

In this context, percent (%) amino acid sequence identity with respect to a particular reference sequence is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in the reference sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity.

- 10 The % identity values used herein may be determined by WU-BLAST-2 which was obtained from [Altschul et al., Methods in Enzymology, 266:460-480 (1996); <http://blast.wustl.edu/blast/README.html>]. WU-BLAST-2 uses several search parameters, most of which are set to the default values. The adjustable
15 parameters are set with the following values: overlap span =1, overlap fraction= 0.125, word threshold (T) = 11. The HSPS and HSPS2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the
20 particular database against which the sequence of interest is being searched; however, the values may be adjusted to increase sensitivity.

A% amino acid sequence identity value is determined by the
25 number of matching identical residues divided by the total number of residues of the "longer" sequence in the aligned region. The "longer" sequence is the one having the most actual residues in the aligned region (gaps introduced by WU-Blast-2 to maximize the alignment score are ignored).
30

In a similar manner, percent (%) nucleic acid sequence identity with respect to a reference nucleic acid sequence is defined as the percentage of nucleotide residues in a candidate sequence that are identical with the nucleotide residues in the reference sequence. The identity values may be determined by the BLASTN module of WU-BLAST-2 set to the
35

default parameters, with overlap span and overlap fraction set to 1 and 0.125, respectively.

Nucleic acid according to the present invention may consist
5 essentially of or consist of the relevant coding sequence.
Nucleic acid according to the present invention may include a
promoter or other regulatory sequence as discussed further
elsewhere herein, and such regulatory sequence may be
heterologous to the coding sequence, that is to say not
10 naturally operably linked therewith. Nucleic acid according
to the present invention may be cDNA or lacking one or more
introns which occur naturally, or may be in any non-naturally
occurring form. A coding sequence in accordance with the
present invention may be included with a larger nucleic acid
15 molecule of less than about 10,000 nucleotides, less than
about 5,000 nucleotides or less than about 2,000 nucleotides.

Also provided by an aspect of the present invention is nucleic
acid including or consisting essentially of a sequence of
20 nucleotides complementary to a nucleotide sequence
hybridisable with any encoding sequence provided herein.
Another way of looking at this would be for nucleic acid
according to this aspect to be hybridisable with a nucleotide
sequence complementary to any encoding sequence provided
25 herein. Of course, DNA is generally double-stranded and
blotting techniques such as Southern hybridisation are often
performed following separation of the strands without a
distinction being drawn between which of the strands is
hybridising. Preferably the hybridisable nucleic acid or its
30 complement encode a product able to influence a physical
characteristic of a plant, particularly vernalization
response, flowering time, leaf shape and/or shade avoidance
response, e.g. in *Arabidopsis thaliana*. Preferred conditions
for hybridisation are familiar to those skilled in the art,
35 but are generally stringent enough for there to be positive
hybridisation between the sequences of interest to the
exclusion of other sequences.

- The nucleic acid, which may contain for example DNA encoding a polypeptide including the amino acid sequence of VRN2 or other polypeptide disclosed herein, as genomic or cDNA, may be in the form of a recombinant and preferably replicable vector,
- 5 for example a plasmid, cosmid, phage or *Agrobacterium* binary vector. The nucleic acid may be under the control of an appropriate promoter or other regulatory elements for expression in a host cell such as a microbial, e.g. bacterial, or plant cell. In the case of genomic DNA, this may contain
- 10 its own promoter or other regulatory elements and in the case of cDNA this may be under the control of an appropriate promoter or other regulatory elements for expression in the host cell.
- 15 A vector including nucleic acid according to the present invention need not include a promoter or other regulatory sequence, particularly if the vector is to be used to introduce the nucleic acid into cells for recombination into the genome.
- 20 Those skilled in the art are well able to construct vectors and design protocols for recombinant gene expression. Suitable vectors can be chosen or constructed, containing appropriate regulatory sequences, including promoter
- 25 sequences, terminator fragments, polyadenylation sequences, enhancer sequences, marker genes and other sequences as appropriate. For further details see, for example, *Molecular Cloning: a Laboratory Manual*: 2nd edition, Sambrook et al., 1989, Cold Spring Harbor Laboratory Press. Many known
- 30 techniques and protocols for manipulation of nucleic acid, for example in preparation of nucleic acid constructs, mutagenesis, sequencing, introduction of DNA into cells and gene expression, and analysis of proteins, are described in detail in *Current Protocols in Molecular Biology*, Second
- 35 Edition, Ausubel et al. eds., John Wiley & Sons, 1992. The disclosures of Sambrook et al. and Ausubel et al. are incorporated herein by reference. Specific procedures and

vectors previously used with wide success upon plants are described by Bevan (Nucl. Acids Res. 12, 8711-8721 (1984)) and Guerineau and Mullineaux (1993) (Plant transformation and expression vectors. In: Plant Molecular Biology Labfax (Croy 5 RRD ed) Oxford, BIOS Scientific Publishers, pp 121-148).

Selectable genetic markers may be used consisting of chimaeric genes that confer selectable phenotypes such as resistance to antibiotics such as kanamycin, hygromycin, phosphinotricin, 10 chlorsulfuron, methotrexate, gentamycin, spectinomycin, imidazolinones and glyphosate.

Nucleic acid molecules and vectors according to the present invention may be provided isolated and/or purified from their 15 natural environment, in substantially pure or homogeneous form, or free or substantially free of nucleic acid or genes of the species of interest or origin other than the sequence encoding a polypeptide with the required function. Nucleic acid according to the present invention may include cDNA, 20 RNA, genomic DNA and may be wholly or partially synthetic. The term "isolate" encompasses all these possibilities. Where a DNA sequence is specified, e.g. with reference to a figure, unless context requires otherwise the RNA equivalent, with U substituted for T where it occurs, is encompassed.

25 When introducing a chosen gene construct into a cell, certain considerations must be taken into account, well known to those skilled in the art. The nucleic acid to be inserted should be assembled within a construct which contains effective 30 regulatory elements which will drive transcription. There must be available a method of transporting the construct into the cell. Once the construct is within the cell membrane, integration into the endogenous chromosomal material either will or will not occur. Finally, as far as plants are 35 concerned the target cell type must be such that cells can be regenerated into whole plants.

Plants transformed with the DNA segment containing the sequence may be produced by standard techniques which are already known for the genetic manipulation of plants. DNA can be transformed into plant cells using any suitable technology, such as a disarmed Ti-plasmid vector carried by *Agrobacterium* exploiting its natural gene transfer ability (EP-A-270355, EP-A-0116718, NAR 12(22) 8711 - 87215 1984), particle or microprojectile bombardment (US 5100792, EP-A-444882, EP-A-434616) microinjection (WO 92/09696, WO 94/00583, EP 331083, EP 175966, Green et al. (1987) *Plant Tissue and Cell Culture*, Academic Press), electroporation (EP 290395, WO 8706614) other forms of direct DNA uptake (DE 4005152, WO 9012096, US 4684611), liposome mediated DNA uptake (e.g. Freeman et al. *Plant Cell Physiol.* 29: 1353 (1984)), or the vortexing method (e.g. Kindle, *PNAS U.S.A.* 87: 1228 (1990d)). Physical methods for the transformation of plant cells are reviewed in Oard, 1991, *Biotech. Adv.* 9: 1-11.

Agrobacterium transformation is widely used by those skilled in the art to transform dicotyledonous species. There are various approaches used for the routine production of stable, fertile transgenic plants in almost all economically relevant monocot plants (Toriyama, et al. (1988) *Bio/Technology* 6, 1072-1074; Zhang, et al. (1988) *Plant Cell Rep.* 7, 379-384; Zhang, et al. (1988) *Theor Appl Genet* 76, 835-840; Shimamoto, et al. (1989) *Nature* 338, 274-276; Datta, et al. (1990) *Bio/Technology* 8, 736-740; Christou, et al. (1991) *Bio/Technology* 9, 957-962; Peng, et al. (1991) International Rice Research Institute, Manila, Philippines 563-574; Cao, et al. (1992) *Plant Cell Rep.* 11, 585-591; Li, et al. (1993) *Plant Cell Rep.* 12, 250-255; Rathore, et al. (1993) *Plant Molecular Biology* 21, 871-884; Fromm, et al. (1990) *Bio/Technology* 8, 833-834; Gordon-Kamm, et al. (1990) *Plant Cell* 2, 603-618; D'Halluin, et al. (1992) *Plant Cell* 4, 1495-1505; Walters, et al. (1992) *Plant Molecular Biology* 18, 189-

200; Koziel, et al. (1993) *Biotechnology* 11, 194-200; Vasil, I. K. (1994) *Plant Molecular Biology* 25, 925-937; Weeks, et al. (1993) *Plant Physiology* 102, 1077-1084; Somers, et al. (1992) *Bio/Technology* 10, 1589-1594; WO92/14828). In particular, *Agrobacterium* mediated transformation is now emerging also as an highly efficient alternative transformation method in monocots (Hiei et al. (1994) *The Plant Journal* 6, 271-282).

10 The generation of fertile transgenic plants has been achieved in the cereals rice, maize, wheat, oat, and barley (reviewed in Shimamoto, K. (1994) *Current Opinion in Biotechnology* 5, 158-162.; Vasil, et al. (1992) *Bio/Technology* 10, 667-674; Vain et al., 1995, *Biotechnology Advances* 13 (4): 653-671;

15 Vasil, 1996, *Nature Biotechnology* 14 page 702).

Microprojectile bombardment, electroporation and direct DNA uptake are preferred where *Agrobacterium* is inefficient or ineffective. Alternatively, a combination of different techniques may be employed to enhance the efficiency of the transformation process, eg bombardment with *Agrobacterium* coated microparticles (EP-A-486234) or microprojectile bombardment to induce wounding followed by co-cultivation with *Agrobacterium* (EP-A-486233).

25 Following transformation, a plant may be regenerated, e.g. from single cells, callus tissue or leaf discs, as is standard in the art. Almost any plant can be entirely regenerated from cells, tissues and organs of the plant. Available techniques 30 are reviewed in Vasil et al., *Cell Culture and Somatic Cell Genetics of Plants, Vol I, II and III, Laboratory Procedures and Their Applications*, Academic Press, 1984, and Weissbach and Weissbach, *Methods for Plant Molecular Biology*, Academic Press, 1989.

The particular choice of a transformation technology will be determined by its efficiency to transform certain plant species as well as the experience and preference of the person practising the invention with a particular methodology of
5 choice. It will be apparent to the skilled person that the particular choice of a transformation system to introduce nucleic acid into plant cells is not essential to or a limitation of the invention, nor is the choice of technique for plant regeneration.

10

A VRN2 gene and modified versions thereof (alleles, mutants, variants and derivatives thereof), and other nucleic acid provided herein, including species homologues, may be used to modify vernalization response, flowering time, leaf shape
15 and/or shade avoidance response in a transgenic plant. Nucleic acid such as a vector as described herein may be used for the production of a transgenic plant. Such a plant may possess an altered phenotype, particular in terms of vernalization response, flowering time, leaf shape and/or
20 shade avoidance response compared with wild-type (that is to say a plant that is wild-type for VRN2 or the relevant homologue thereof).

The invention further encompasses a host cell transformed with
25 nucleic acid or a vector according to the present invention, especially a plant or a microbial cell. Thus, a host cell, such as a plant cell, including heterologous nucleic acid according to the present invention is provided. Within the cell, the nucleic acid may be incorporated within the
30 chromosome. There may be more than one heterologous nucleotide sequence per haploid genome.

Also according to the invention there is provided a plant cell having incorporated into its genome nucleic acid, particularly
35 heterologous nucleic acid, as provided by the present invention, under operative control of a regulatory sequence for control of expression. The coding sequence may be operably linked to one or more regulatory sequences which may

be heterologous or foreign to the gene, such as not naturally associated with the gene for its expression. The nucleic acid according to the invention may be placed under the control of an externally inducible gene promoter to place expression
5 under the control of the user.

A suitable inducible promoter is the GST-II-27 gene promoter which has been shown to be induced by certain chemical compounds which can be applied to growing plants. The
10 promoter is functional in both monocotyledons and dicotyledons. It can therefore be used to control gene expression in a variety of genetically modified plants, including field crops such as canola, sunflower, tobacco, sugarbeet, cotton; cereals such as wheat, barley, rice, maize,
15 sorghum; fruit such as tomatoes, mangoes, peaches, apples, pears, strawberries, bananas, and melons; and vegetables such as carrot, lettuce, cabbage and onion. The GST-II-27 promoter is also suitable for use in a variety of tissues, including roots, leaves, stems and reproductive tissues.
20

A further aspect of the present invention provides a method of making such a plant cell involving introduction of nucleic acid or a suitable vector including the sequence of nucleotides into a plant cell and causing or allowing
25 recombination between the vector and the plant cell genome to introduce the sequence of nucleotides into the genome. The invention extends to plant cells containing nucleic acid according to the invention as a result of introduction of the nucleic acid into an ancestor cell.

30 The term "heterologous" may be used to indicate that the gene/sequence of nucleotides in question have been introduced into said cells of the plant or an ancestor thereof, using genetic engineering, ie by human intervention. A transgenic plant cell, i.e. transgenic for the nucleic acid in question, may be provided. The transgene may be on an extra-genomic vector or incorporated, preferably stably, into the genome. A heterologous gene may replace an endogenous equivalent gene,
35

ie one which normally performs the same or a similar function, or the inserted sequence may be additional to the endogenous gene or other sequence. An advantage of introduction of a heterologous gene is the ability to place expression of a sequence under the control of a promoter of choice, in order to be able to influence expression according to preference. Furthermore, mutants, variants and derivatives of the wild-type gene, e.g. with higher or lower activity than wild-type, may be used in place of the endogenous gene. Nucleic acid heterologous, or exogenous or foreign, to a plant cell may be non-naturally occurring in cells of that type, variety or species. Thus, nucleic acid may include a coding sequence of or derived from a particular type of plant cell or species or variety of plant, placed within the context of a plant cell of a different type or species or variety of plant. A further possibility is for a nucleic acid sequence to be placed within a cell in which it or a homologue is found naturally, but wherein the nucleic acid sequence is linked and/or adjacent to nucleic acid which does not occur naturally within the cell, or cells of that type or species or variety of plant, such as operably linked to one or more regulatory sequences, such as a promoter sequence, for control of expression. A sequence within a plant or other host cell may be identifiably heterologous, exogenous or foreign.

Plants which include a plant cell according to the invention are also provided, along with any part or propagule thereof, seed, selfed or hybrid progeny and descendants. A plant according to the present invention may be one which does not breed true in one or more properties. Plant varieties may be excluded, particularly registrable plant varieties according to Plant Breeders' Rights. It is noted that a plant need not be considered a "plant variety" simply because it contains stably within its genome a transgene, introduced into a cell of the plant or an ancestor thereof.

In addition to a plant, the present invention provides any clone of such a plant, seed, selfed or hybrid progeny and

- descendants, and any part of any of these, such as cuttings, seed. The invention provides any plant propagule, that is any part which may be used in reproduction or propagation, sexual or asexual, including cuttings, seed and so on. Also
- 5 encompassed by the invention is a plant which is a sexually or asexually propagated off-spring, clone or descendant of such a plant, or any part or propagule of said plant, off-spring, clone or descendant.
- 10 The invention further provides a method of influencing or affecting a physical characteristic of a plant, particularly vernalization response, flowering time, leaf shape and/or shade avoidance response, including causing or allowing expression of a heterologous nucleic acid sequence as
- 15 discussed within cells of the plant.
- The invention further provides a method of inducing expression from nucleic acid encoding a VRN2 polypeptide, or a mutant, variant, allele or derivative of the sequence, or a homologue,
- 20 according to the disclosure herein, within cells of a plant (thereby producing the encoded polypeptide), following an earlier step of introduction of the nucleic acid into a cell of the plant or an ancestor thereof. Such a method may influence or affect a characteristic of the plant, such as
- 25 vernalization response, flowering time, leaf shape and/or shade avoidance response. This may be used in combination with any other gene, such as transgenes involved in flowering (e.g. FCA) or other phenotypic trait or desirable property.
- 30 The present invention also encompasses the expression product of any of the nucleic acid sequences disclosed and methods of making the expression product by expression from encoding nucleic acid therefore under suitable conditions, which may be in suitable host cells. Following expression, the product may
- 35 be isolated from the expression system and may be used as desired, for instance in formulation of a composition including at least one additional component.

The present invention also provides for the production and use of fragments of the full-length polypeptides disclosed herein, especially active portions thereof. An "active portion" of a polypeptide means a peptide which is less than said full length polypeptide, but which retains an essential biological activity. In particular, the active portion retains the ability to alter vernalization response, flowering time, leaf shape and/or shade avoidance response in a plant, such as *Arabidopsis thaliana*.

10

A "fragment" of a polypeptide means a stretch of amino acid residues of at least about five to seven contiguous amino acids, often at least about seven to nine contiguous amino acids, typically at least about nine to 13 contiguous amino acids and, most preferably, at least about 20 to 30 or more contiguous amino acids. Fragments of the polypeptides may include one or more epitopes useful for raising antibodies to a portion of any of the amino acid sequences disclosed herein. Preferred epitopes are those to which antibodies are able to bind specifically, which may be taken to be binding a polypeptide or fragment thereof of the invention with an affinity which is at least about 1000x that of other polypeptides.

25 Among preferred VRN2 fragments according to the present invention are the zinc finger domain, DNA binding domain and other domains disclosed herein.

Purified protein according to the present invention, or a 30 fragment, mutant, derivative or variant thereof, e.g. produced recombinantly by expression from encoding nucleic acid therefor, may be used to raise antibodies employing techniques which are standard in the art. Antibodies and polypeptides comprising antigen-binding fragments of antibodies may be used 35 in identifying homologues from other species as discussed further below, also in identifying complexes containing VRN2 protein.

Methods of producing antibodies include immunising a mammal (e.g. human, mouse, rat, rabbit, horse, goat, sheep or monkey) with the protein or a fragment thereof. Antibodies may be obtained from immunised animals using any of a variety of 5 techniques known in the art, and might be screened, preferably using binding of antibody to antigen of interest. For instance, Western blotting techniques or immunoprecipitation may be used (Armitage et al, 1992, Nature 357: 80-82).

Antibodies may be polyclonal or monoclonal.

10

As an alternative or supplement to immunising a mammal, antibodies with appropriate binding specificity may be obtained from a recombinantly produced library of expressed immunoglobulin variable domains, e.g. using lambda 15 bacteriophage or filamentous bacteriophage which display functional immunoglobulin binding domains on their surfaces; for instance see WO92/01047.

Antibodies raised to a polypeptide or peptide can be used in 20 the identification and/or isolation of homologous polypeptides, and then the encoding genes. Thus, the present invention provides a method of identifying or isolating a polypeptide with the desired function (in accordance with embodiments disclosed herein), comprising screening candidate 25 polypeptides with a polypeptide comprising the antigen-binding domain of an antibody (for example whole antibody or a suitable fragment thereof, e.g. scFv, Fab) which is able to bind a polypeptide or fragment, variant or derivative thereof according to the present invention or preferably has binding 30 specificity for such a polypeptide. Specific binding members such as antibodies and polypeptides comprising antigen binding domains of antibodies that bind and are preferably specific for a polypeptide or mutant, variant or derivative thereof according to the invention represent further aspects of the 35 present invention, particularly in isolated and/or purified form, as do their use and methods which employ them.

Candidate polypeptides for screening may for instance be the products of an expression library created using nucleic acid derived from an plant of interest, or may be the product of a purification process from a natural source. A polypeptide 5 found to bind the antibody may be isolated and then may be subject to amino acid sequencing. Any suitable technique may be used to sequence the polypeptide either wholly or partially (for instance a fragment of the polypeptide may be sequenced). Amino acid sequence information may be used in obtaining 10 nucleic acid encoding the polypeptide, for instance by designing one or more oligonucleotides (e.g. a degenerate pool of oligonucleotides) for use as probes or primers in hybridization to candidate nucleic acid, or by searching computer sequence databases, as discussed further below.

15

A further aspect of the present invention provides a method of identifying and cloning VRN2 homologues from plant species other than *Arabidopsis thaliana* which method employs a nucleotide sequence derived from that shown herein. As 20 discussed above, sequences derived from these may themselves be used in identifying and in cloning other sequences. The nucleotide sequence information provided herein, or any part thereof, may be used in a data-base search to find homologous sequences, expression products of which can be tested for 25 ability to influence a plant characteristic. These may have ability to alter vernalization response, flowering time, leaf shape and/or shade avoidance response in a plant. Alternatively, nucleic acid libraries may be screened using techniques well known to those skilled in the art and 30 homologous sequences thereby identified then tested.

The present invention also extends to nucleic acid encoding a VRN2 homologue obtained using a nucleotide sequence derived from any of those shown herein.

35

In certain embodiments, nucleic acid according to the present invention encodes a polypeptide which has homology with all or part of VRN2 amino acid sequence shown herein, in the terms

- discussed already above (e.g. for length), which homology is greater over the length of the relevant part (i.e. fragment) than the homology shared between a respective part of the VRN2 amino acid sequence of *Arabidopsis thaliana*, and the other sequences shown in Figure 8a or Figure 8b, and may be greater than about 5% greater, more preferably greater than about 10% greater, more preferably greater than about 20% greater, and more preferably greater than about 30% greater. Thus, to exemplify with reference to one embodiment, nucleic acid encoding an amino acid mutant, variant or derivative of the amino acid sequence shown in SEQ ID NO: 2 may be provided wherein the encoded amino acid sequence includes a contiguous sequence of about 100 amino acids which has greater homology with a contiguous sequence of 100 amino acids within the amino acid sequence of SEQ ID NO: 2 than any contiguous sequence of 100 amino acids within another sequence shown in Figure 8a or 8b, preferably greater than about 5% greater homology, and so on.
- Similarly, nucleic acid according to certain embodiments of the present invention may have homology with all or part of a nucleotide sequence shown herein, in the terms discussed already above (e.g. for length), which homology is greater over the length of the relevant part (i.e. fragment) than the homology shared with a respective part of the natural coding nucleotide sequence for the other amino acid sequences shown in Figure 8a or 8b and referenced herein, and may be greater than about 5% greater, more preferably greater than about 10% greater, more preferably greater than about 20% greater, and more preferably greater than about 30% greater.

The provision of sequence information for the VRN2 gene of *Arabidopsis thaliana* enables the obtention of homologous sequences from other plant species. In particular, homologues may be easily isolated from related, commercially important species that have a vernalization requirement, or show some response to vernalization. These would include all members of the Brassicaceae, and other dicots including

tobacco, sugarbeet, peas and celery. Monocots included in this category are the cereals rice, wheat and barley.

Thus, included within the scope of the present invention are 5 nucleic acid molecules which encode amino acid sequences which are homologues of VRN2 of *Arabidopsis thaliana*. Homology may be at the nucleotide sequence and/or amino acid sequence level, as has already been discussed above. A homologue from a species other than *Arabidopsis thaliana* encodes a product 10 which causes a phenotype similar to that caused by the VRN2 gene, generally including ability to alter vernalization response, flowering time, leaf shape and/or shade avoidance response in a plant, such as in *Arabidopsis thaliana*. In addition, mutants, derivatives or alleles of these genes may 15 have altered, e.g. increased or decreased, activity or ability compared with wild-type.

VRN2 gene homologues may also be identified from economically important monocotyledonous crop plants including the cereals 20 rice, wheat and barley. Although genes encoding the same protein in monocotyledonous and dicotyledonous plants show relatively little homology at the nucleotide level, amino acid sequences are conserved. Therefore it is possible to use public sequence databases to identify *Arabidopsis*, rice or 25 maize cDNA clone sequences that were obtained in random sequencing programmes and share homology to the gene of interest, as has been done for other genes isolated from *Arabidopsis* (e.g CO; WO 96/14414). Of course, mutants, derivatives and alleles of these sequences are included within 30 the scope of the present invention in the same terms as discussed above for the *Arabidopsis thaliana* VRN2 gene.

According to a further aspect, the present invention provides a method of identifying or a method of cloning a VRN2 35 homologue, e.g. from a species other than *Arabidopsis thaliana* the method employing a nucleotide sequence derived from any of those shown herein. For instance, such a method may employ an

oligonucleotide or oligonucleotides which comprises or consists of a sequence or sequences conserved between or encoding a sequence or sequences conserved between the sequences shown in Figure 8a or 8b, or a sequence or sequences 5 conserved between the sequences of SEQ ID NO: 2 and SEQ ID NO: 5, or encoding sequences SEQ ID NO: 1 and SEQ ID NO: 4, to search for homologues. Thus, a method of obtaining nucleic acid is provided, comprising hybridisation of an oligonucleotide or a nucleic acid molecule comprising such an 10 oligonucleotide to target/candidate nucleic acid. Target or candidate nucleic acid may, for example, comprise a genomic or cDNA library obtainable from an organism known to contain or suspected of containing such nucleic acid, either monocotyledonous or dicotyledonous. Successful hybridisation 15 may be identified and target/candidate nucleic acid isolated for further investigation and/or use.

Hybridisation may involve probing nucleic acid and identifying positive hybridisation under suitably stringent conditions (in 20 accordance with known techniques) and/or use of oligonucleotides as primers in a method of nucleic acid amplification, such as PCR. For probing, preferred conditions are those which are stringent enough for there to be a simple pattern with a small number of hybridisations identified as 25 positive which can be investigated further. It is well known in the art to increase stringency of hybridisation gradually until only a few positive clones remain.

For instance, screening may initially be carried out under 30 conditions, which comprise a temperature of about 37°C or more, a formamide concentration of less than about 50%, and a moderate to low salt (e.g. Standard Saline Citrate ('SSC') = 0.15 M sodium chloride; 0.15 M sodium citrate; pH 7) concentration.

35

Alternatively, a temperature of about 50°C or more and a high salt (e.g. 'SSPE'= 0.180 M sodium chloride; 9 mM disodium hydrogen phosphate; 9 mM sodium dihydrogen phosphate; 1 mM

sodium EDTA; pH 7.4). Preferably the screening is carried out at about 37°C, a formamide concentration of about 20%, and a salt concentration of about 5 X SSC, or a temperature of about 50°C and a salt concentration of about 2 X SSPE. These

5 conditions will allow the identification of sequences which have a substantial degree of homology (similarity, identity) with the probe sequence, without requiring the perfect homology for the identification of a stable hybrid.

10 Suitable conditions include, e.g. for detection of sequences that are about 80-90% identical, hybridization overnight at 42°C in 0.25M Na₂HPO₄, pH 7.2, 6.5% SDS, 10% dextran sulfate and a final wash at 55°C in 0.1X SSC, 0.1% SDS. For detection of sequences that are greater than about 90% identical,

15 suitable conditions include hybridization overnight at 65°C in 0.25M Na₂HPO₄, pH 7.2, 6.5% SDS, 10% dextran sulfate and a final wash at 60°C in 0.1X SSC, 0.1% SDS.

An alternative is a solution of 5x SSPE (final 0.9 M NaCl,
20 0.05M sodium phosphate, 0.005M ethylenediaminetetraacetic acid (EDTA) pH 7.7), 5X Denhardt's solution, 0.5% SDS (sodium dodecyl sulphate), at 65°C overnight, (for high stringency, highly similar sequences) or 50°C (for low stringency, less similar sequences). Washes in 0.2x SSC/0.1% SDS at 65°C for
25 high stringency, alternatively at 50-60°C in 1x SSC/0.1% SDS for low stringency.

The present invention extends to nucleic acid selectively hybridisable under high stringency with nucleic acid
30 identified herein.

As an alternative to probing, though still employing nucleic acid hybridisation, oligonucleotides designed to amplify DNA sequences may be used in PCR reactions or other methods
35 involving amplification of nucleic acid, using routine procedures. See for instance "PCR protocols; A Guide to Methods and Applications", Eds. Innis et al, 1990, Academic Press, New York.

Preferred amino acid sequences suitable for use in the design of probes or PCR primers for some purposes are sequences conserved (completely, substantially or partly) between the VRN2 sequence and at least one other of the sequences shown in 5 Figure 8a or 8b.

Preferred primers for amplification of conserved regions of VRN2 for use as probes to obtain genomic or cDNA clones may include the following:

10 Primers VRN2-AI and VRN2-AJ which, in RT-PCR, amplify a 1583 bp fragment that contains the complete VRN2 open reading frame, and portions of both the 5' and 3' untranslated sequences;

15 Primers VRN2-AP and VRN2-AJ which, in RT-PCR, amplify a 781 bp fragment that includes the conserved acidic region;

20 Primers VRN2-AO and VRN2-AS which, in RT-PCR, amplify a 493 bp fragment that includes the zinc-finger motif, and the second nuclear localization signal (NLS); and

25 Primers VRN2-AI and VRN2-AJ from genomic DNA, which amplify a 3605 bp product that includes most of the VRN2 gene, except the promoter and 3' regions (i.e. encompasses the same regions as the VRN2-AI/AJ pair above, but with the introns, useful for hybridisation to genomic DNA, less so for cDNA).

On the basis of amino acid sequence information
30 oligonucleotide probes or primers may be designed, taking into account the degeneracy of the genetic code, and, where appropriate, codon usage of the organism from which the candidate nucleic acid is derived.

35 Preferably an oligonucleotide in accordance with certain embodiments of the invention, e.g. for use in nucleic acid amplification, is up to about 50 nucleotides, or about 40

nucleotides or about 30 or fewer nucleotides in length (e.g. 18, 21 or 24).

Assessment of whether or not such a PCR product corresponds to
5 a homologue gene may be conducted in various ways. A PCR band
from such a reaction might contain a complex mix of products.
Individual products may be cloned and each one individually
screened. It may be analysed by transformation to assess
function on introduction into a plant of interest.

10

As noted, nucleic acid according to the present invention is obtainable using oligonucleotides, designed on the basis of sequence information provided herein, as probes or primers. Nucleic acid isolated and/or purified from one or more cells
15 of a plant (see above), or a nucleic acid library derived from nucleic acid isolated and/or purified from the plant (e.g. a cDNA library derived from mRNA isolated from the plant), may be probed under conditions for selective hybridisation and/or subjected to a specific nucleic acid amplification reaction
20 such as the polymerase chain reaction (PCR). The nucleic acid probed or used as template in the amplification reaction may be genomic DNA, cDNA or RNA. If necessary, one or more gene fragments may be ligated to generate a full-length coding sequence.

25

PCR primers derived from the VRN2 sequences disclosed herein may readily be tested for their specificity for amplifying nucleic acid according to the present invention, using both genomic DNA and RT-PCR templates. Cloning and subsequent
30 sequencing of PCR products may be used to indicate amplification of the expected derived gene fragment. Full length cDNA clones can be obtained as described by 5' and 3' RACE technology if RT-PCR products are used as templates.

35 Various aspects of the present invention include the obtainable nucleic acid, methods of screening material, e.g. cell lysate, nucleic acid preparations, for the presence of

nucleic acid of interest, methods of obtaining the nucleic acid, and suitable primers and primer combinations.

The sequence information provided herein also allows the
5 design of diagnostic tests for determination of the presence
of a specific gene or allele thereof in any given plant,
cultivar, variety, population, landrace, part of a family or
other selection in a breeding programme or other such
genotype. A diagnostic test may be based on determination of
10 the presence or absence of a particular allele by means of
nucleic acid or polypeptide determination.

At the nucleic acid level, this may involve hybridisation of a
suitable oligo- or poly-nucleotide, such as a fragment of the
15 gene or a homologue thereof, including any homologue disclosed
herein, or any particular allele, such as an allele which
gives a desirable phenotype, such as any such allele disclosed
herein. The hybridisation may involve PCR designed to amplify
a product from a given allelic version of the gene, with
20 subsequent detection of an amplified product by any of a
number of possible methods including but not limited to gel
electrophoresis, capillary electrophoresis, direct
hybridisation of nucleotide sequence probes and so on. A
diagnostic test may be based on PCR designed to amplify
25 various alleles or any allele from the relevant locus, with a
test to distinguish the different possible alleles by any of a
number of possible methods, including DNA fragment size,
restriction site variation (e.g. CAPS - cleaved amplified
polymorphic sites) and so on. A diagnostic test may also be
30 based on a great number of possible variants of nucleic acid
analysis that will be apparent to those skilled in the art,
such as use of a synthetic sequence as a hybridisation probe.

Broadly, the methods divide into those screening for the
35 presence of nucleic acid sequences and those that rely on
detecting the presence or absence of a polypeptide. The
methods may make use of biological samples from one or more

plants or cells that are suspected to contain the nucleic acid sequences or polypeptide.

Exemplary approaches for detecting nucleic acid or 5 polypeptides include analysing a sample from the plant or plant cell by:

- (a) comparing the sequence of nucleic acid in the sample with all or part of a nucleotide sequence shown herein, to 10 determine whether the sample contains a mutation;
- (b) determining the presence in the sample of a polypeptide including a VRN2 amino acid sequence shown herein, or a fragment thereof and, if present, determining whether the 15 polypeptide is full length, and/or is mutated, and/or is expressed at the normal level;
- (c) performing DNA fingerprinting to compare the restriction pattern produced when a restriction enzyme cuts nucleic acid 20 in the sample with the restriction pattern obtained from a nucleotide sequence shown herein, or from a known mutant, allele or variant thereof;
- (d) contacting the sample with a specific binding member 25 capable of binding to nucleic acid including the nucleotide sequence as set out herein, or a fragment thereof, or a mutant, allele or variant thereof, the specific binding member including nucleic acid hybridisable with a VRN2 sequence herein, or a polypeptide including a binding domain with 30 specificity for nucleic acid including a VRN2 sequence or polypeptide encoded by it, or a mutated form thereof, and determining binding of the specific binding member;
- (e) performing PCR involving one or more primers based on a 35 nucleotide sequence shown herein to screen the sample for nucleic acid including the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO: 4 or a mutant, allele or variant thereof.

When screening for a VRN2 allele nucleic acid, the nucleic acid in the sample will initially be amplified, e.g. using PCR, to increase the amount of the analyte as compared to other sequences present in the sample. This allows the target 5 sequences to be detected with a high degree of sensitivity if they are present in the sample. This initial step may be avoided by using highly sensitive array techniques that are becoming increasingly important in the art.

10 A variant form of the gene may contain one or more insertions, deletions, substitutions and/or additions of one or more nucleotides compared with the wild-type sequence which may or may not disrupt or alter the gene function. Differences at the nucleic acid level are not necessarily reflected by a 15 difference in the amino acid sequence of the encoded polypeptide. However, a mutation or other difference in a gene may result in a frame-shift or stop codon, which could seriously affect the nature of the polypeptide produced (if any), or a point mutation or gross mutational change to the 20 encoded polypeptide, including insertion, deletion, substitution and/or addition of one or more amino acids or regions in the polypeptide. A mutation in a promoter sequence or other regulatory region may prevent or reduce expression from the gene or affect the processing or stability of the 25 mRNA transcript.

Tests may be carried out on preparations containing genomic DNA, cDNA and/or mRNA. Testing cDNA or mRNA has the advantage of the complexity of the nucleic acid being reduced by the 30 absence of intron sequences, but the possible disadvantage of extra time and effort being required in making the preparations. RNA is more difficult to manipulate than DNA because of the wide-spread occurrence of RN'ases.

35 Nucleic acid in a test sample may be sequenced and the sequence compared with a sequence shown herein to determine whether or not a difference is present. If so, the difference can be compared with known alleles to determine whether the

test nucleic acid contains one or more of the variations indicated, or the difference can be investigated for association with a desired phenotype.

- 5 The amplified nucleic acid may then be sequenced as above, and/or tested in any other way to determine the presence or absence of a particular feature. Nucleic acid for testing may be prepared from nucleic acid removed from cells or in a library using a variety of other techniques such as
10 restriction enzyme digest and electrophoresis.

Nucleic acid may be screened using a variant- or allele-specific probe. Such a probe corresponds in sequence to a region of the gene, or its complement, containing a sequence
15 alteration known to be associated with alteration of ability to affect vernalization response, flowering time, leaf shape and/or shade avoidance response. Under suitably stringent conditions, specific hybridisation of such a probe to test nucleic acid is indicative of the presence of the sequence
20 alteration in the test nucleic acid. For efficient screening purposes, more than one probe may be used on the same test sample.

Allele- or variant-specific oligonucleotides may similarly be
25 used in PCR to specifically amplify particular sequences if present in a test sample. Assessment of whether a PCR band contains a gene variant may be carried out in a number of ways familiar to those skilled in the art. The PCR product may for instance be treated in a way that enables one to display the
30 mutation or polymorphism on a denaturing polyacrylamide DNA sequencing gel, with specific bands that are linked to the gene variants being selected.

An alternative or supplement to looking for the presence of
35 variant sequences in a test sample is to look for the presence of the normal sequence, e.g. using a suitably specific oligonucleotide probe or primer.

Approaches which rely on hybridisation between a probe and test nucleic acid and subsequent detection of a mismatch may be employed. Under appropriate conditions (temperature, pH etc.), an oligonucleotide probe will hybridise with a sequence 5 which is not entirely complementary. The degree of base-pairing between the two molecules will be sufficient for them to anneal despite a mis-match. Various approaches are well known in the art for detecting the presence of a mis-match between two annealing nucleic acid molecules.

10

For instance, RN'ase A cleaves at the site of a mis-match. Cleavage can be detected by electrophoresing test nucleic acid to which the relevant probe or probe has annealed and looking for smaller molecules (i.e. molecules with higher 15 electrophoretic mobility) than the full length probe/test hybrid. Other approaches rely on the use of enzymes such as resolvases or endonucleases.

Thus, an oligonucleotide probe that has the sequence of a 20 region of the normal gene (either sense or anti-sense strand) in which mutations associated with particular phenotypes are known to occur may be annealed to test nucleic acid and the presence or absence of a mis-match determined. Detection of the presence of a mis-match may indicate the presence in the 25 test nucleic acid of a mutation. On the other hand, an oligonucleotide probe that has the sequence of a region of the gene including a mutation may be annealed to test nucleic acid and the presence or absence of a mis-match determined. The presence of a mis-match may indicate that the nucleic acid in 30 the test sample has the normal sequence, or a different mutant or allele sequence. In either case, a battery of probes to different regions of the gene may be employed.

The presence of differences in sequence of nucleic acid 35 molecules may be detected by means of restriction enzyme digestion, such as in a method of DNA fingerprinting where the restriction pattern produced when one or more restriction enzymes are used to cut a sample of nucleic acid is compared

with the pattern obtained when a sample containing the normal gene or a variant or allele is digested with the same enzyme or enzymes.

- 5 The presence of absence of a lesion in a promoter or other regulatory sequence may also be assessed by determining the level of mRNA production by transcription or the level of polypeptide production by translation from the mRNA.
- 10 Nucleic acid isolated and/or purified from one or more cells of a plant or a nucleic acid library derived from nucleic acid isolated and/or purified from cells (e.g. a cDNA library derived from mRNA isolated from the cells), may be probed under conditions for selective hybridisation and/or subjected
- 15 to a specific nucleic acid amplification reaction such as the polymerase chain reaction (PCR).

A method may include hybridisation of one or more (e.g. two) probes or primers to target nucleic acid. Where the nucleic acid is double-stranded DNA, hybridisation will generally be preceded by denaturation to produce single-stranded DNA. The hybridisation may be as part of a PCR procedure, or as part of a probing procedure not involving PCR. An example procedure would be a combination of PCR and low stringency hybridisation. A screening procedure, chosen from the many available to those skilled in the art, is used to identify successful hybridisation events and isolate hybridised nucleic acid.

- 30 Binding of a probe to target nucleic acid (e.g. DNA) may be measured using any of a variety of techniques at the disposal of those skilled in the art. For instance, probes may be radioactively, fluorescently or enzymatically labelled. Other methods not employing labelling of probe include examination
- 35 of restriction fragment length polymorphisms, amplification using PCR, RNAase cleavage and allele specific oligonucleotide probing.

Probing may employ the standard Southern blotting technique. For instance DNA may be extracted from cells and digested with different restriction enzymes. Restriction fragments may then be separated by electrophoresis on an agarose gel, before

5 denaturation and transfer to a nitrocellulose filter.

Labelled probe may be hybridised to the DNA fragments on the filter and binding determined. DNA for probing may be prepared from RNA preparations from cells.

- 10 Preliminary experiments may be performed by hybridising under low stringency conditions various probes to Southern blots of DNA digested with restriction enzymes. Suitable conditions would be achieved when a large number of hybridising fragments were obtained while the background hybridisation was low.
- 15 Using these conditions nucleic acid libraries, e.g. cDNA libraries representative of expressed sequences, may be searched.

As noted, those skilled in the art are well able to employ
20 suitable conditions of the desired stringency for selective hybridisation, taking into account factors such as oligonucleotide length and base composition, temperature and so on.

25 In some preferred embodiments of diagnostic assays according to the present invention, oligonucleotides according to the present invention that are fragments of any of the sequences shown herein, or any allele associated with a desired phenotype are at least about 10 nucleotides in length, more
30 preferably at least about 15 nucleotides in length, more preferably at least about 20 nucleotides in length, more preferably about 30 nucleotides in length. Such fragments themselves individually represent aspects of the present invention. Fragments and other oligonucleotides may be used
35 as primers or probes as discussed but may also be generated (e.g. by PCR) in methods concerned with determining the presence in a test sample of a sequence indicative of a desired phenotype.

There are various methods for determining the presence or absence in a test sample of a particular polypeptide, such as a polypeptide including the amino acid sequence shown in SEQ

5 ID NO: 2 or SEQ ID NO: 5, or an amino acid sequence mutant, variant or allele thereof.

A sample may be tested for the presence of a binding partner for a specific binding member such as an antibody (or mixture 10 of antibodies), specific for one or more particular variants of a polypeptide shown herein.

In such cases, the sample may be tested by being contacted with a specific binding member such as an antibody under 15 appropriate conditions for specific binding, before binding is determined, for instance using a reporter system as discussed. Where a panel of antibodies is used, different reporting labels may be employed for each antibody so that binding of each can be determined.

20 A specific binding member such as an antibody may be used to isolate and/or purify its binding partner polypeptide from a test sample, to allow for sequence and/or biochemical analysis of the polypeptide to determine whether it has the sequence 25 and/or properties of the wild-type polypeptide or a particular mutant, variant or allele thereof. Amino acid sequence is routine in the art using automated sequencing machines.

The use of diagnostic tests for alleles allows the researcher 30 or plant breeder to establish, with full confidence and independent from time consuming biochemical tests, whether or not a desired allele is present in the plant of interest (or a cell thereof), whether the plant is a representative of a collection of other genetically identical plants (e.g. an 35 inbred variety or cultivar) or one individual in a sample of related (e.g. breeders' selection) or unrelated plants.

In a breeding scheme based on selection and selfing of desirable individuals, nucleic acid or polypeptide diagnostics for the desirable allele or alleles in high throughput, low cost assays as provided by this invention, reliable selection 5 for the can be made at early generations and on more material than would otherwise be possible. This gain in reliability of selection plus the time saving by being able to test material earlier and without costly phenotype screening is of considerable value in plant breeding.

10

Nucleic acid-based determination of the presence or absence of one or more desirable alleles may be combined with determination of the genotype of the flanking linked genomic DNA and other unlinked genomic DNA using established sets of 15 markers such as RFLPs, microsatellites or SSRs, AFLPs, RAPDs etc. This enables the researcher or plant breeder to select for not only the presence of the desirable allele but also for individual plant or families of plants which have the most desirable combinations of linked and unlinked genetic 20 background. Such recombinations of desirable material may occur only rarely within a given segregating breeding population or backcross progeny. Direct assay of the locus as afforded by the present invention allows the researcher to make a stepwise approach to fixing (making homozygous) the 25 desired combination of flanking markers and alleles, by first identifying individuals fixed for one flanking marker and then identifying progeny fixed on the other side of the locus all the time knowing with confidence that the desirable allele is still present.

30

The present disclosure provides sufficient information for a person skilled in the art to obtain genomic DNA sequence for any given new or existing allele and devise a suitable nucleic acid- and/or polypeptide-based diagnostic assay. In designing 35 a nucleic acid assay account is taken of the distinctive variation in sequence that characterises the particular variant allele.

Nucleic acid according to the invention may include a nucleotide sequence encoding a product involved in vernalization response, flowering time, leaf shape and/or shade avoidance response. Reducing or increasing the level of 5 expression may be used to manipulate such a characteristic in a plant. This may involve use of anti-sense or sense regulation, discussed further below.

Nucleic acid according to the invention, such as a VRN2 gene 10 or homologue, may be placed under the control of an externally inducible gene promoter to place expression under the control of the user. An advantage of introduction of a heterologous gene into a plant cell, particularly when the cell is comprised in a plant, is the ability to place expression of 15 the gene under the control of a promoter of choice, in order to be able to influence gene expression, and therefore vernalization response, flowering time, leaf shape, shade avoidance response, and/or other characteristic, according to preference. Furthermore, mutants and derivatives of the 20 wild-type gene, e.g. with higher or lower activity than wild-type, may be used in place of the endogenous gene.

In the present invention, over-expression may be achieved by 25 introduction of the nucleotide sequence in a sense orientation. Thus, the present invention provides a method of influencing a physical characteristic of a plant, the method including causing or allowing expression of the product (polypeptide or nucleic acid transcript) encoded by heterologous nucleic acid according to the invention from that 30 nucleic acid within cells of the plant.

Down-regulation of expression of a target gene may be achieved using anti-sense technology or "sense regulation" ("co-suppression").

35

In using anti-sense genes or partial gene sequences to down-regulate gene expression, a nucleotide sequence is placed under the control of a promoter in a "reverse orientation"

- such that transcription yields RNA which is complementary to normal mRNA transcribed from the "sense" strand of the target gene. See, for example, Rothstein et al, 1987; Smith et al, (1988) *Nature* 334, 724-726; Zhang et al, (1992) *The Plant Cell* 4, 1575-1588, English et al., (1996) *The Plant Cell* 8, 179-188. Antisense technology is also reviewed in Bourque, (1995), *Plant Science* 105, 125-149, and Flavell, (1994) *PNAS USA* 91, 3490-3496.
- 10 An alternative is to use a copy of all or part of the target gene inserted in sense, that is the same orientation as the target gene, to achieve reduction in expression of the target gene by co-suppression. See, for example, van der Krol et al., (1990) *The Plant Cell* 2, 291-299; Napoli et al., (1990) *The Plant Cell* 2, 279-289; Zhang et al., (1992) *The Plant Cell* 4, 1575-1588, and US-A-5,231,020.

20 The complete sequence corresponding to the coding sequence (in reverse orientation for anti-sense) need not be used. For example fragments of sufficient length may be used. It is a routine matter for the person skilled in the art to screen fragments of various sizes and from various parts of the coding sequence to optimise the level of anti-sense inhibition. It may be advantageous to include the initiating 25 methionine ATG codon, and perhaps one or more nucleotides upstream of the initiating codon. A further possibility is to target a conserved sequence of a gene, e.g. a sequence that is characteristic of one or more genes, such as a regulatory sequence.

30 The sequence employed may be about 500 nucleotides or less, possibly about 400 nucleotides, about 300 nucleotides, about 200 nucleotides, or about 100 nucleotides. It may be possible to use oligonucleotides of much shorter lengths, 14-23 35 nucleotides, although longer fragments, and generally even longer than about 500 nucleotides are preferable where possible, such as longer than about 600 nucleotides, than

about 700 nucleotides, than about 800 nucleotides, than about 1000 nucleotides or more.

It may be preferable that there is complete sequence identity
5 in the sequence used for down-regulation of expression of a target sequence, and the target sequence, though total complementarity or similarity of sequence is not essential. One or more nucleotides may differ in the sequence used from the target gene. Thus, a sequence employed in a down-
10 regulation of gene expression in accordance with the present invention may be a wild-type sequence (e.g. gene) selected from those available, or a mutant, derivative, variant or allele, by way of insertion, addition, deletion or substitution of one or more nucleotides, of such a sequence.
15 The sequence need not include an open reading frame or specify an RNA that would be translatable. It may be preferred for there to be sufficient homology for the respective anti-sense and sense RNA molecules to hybridise. There may be down regulation of gene expression even where there is about 5%,
20 10%, 15% or 20% or more mismatch between the sequence used and the target gene.

Generally, the transcribed nucleic acid may represent a fragment of a gene, or the complement thereof, or may be a
25 mutant, derivative, variant or allele thereof, in similar terms as discussed above in relation to alterations being made to a coding sequence and the homology of the altered sequence. The homology may be sufficient for the transcribed anti-sense RNA to hybridise with nucleic acid within cells of the plant,
30 though irrespective of whether hybridisation takes place the desired effect is down-regulation of gene expression.

Thus, the present invention also provides a method of modifying, affecting, altering or modulating a characteristic
35 of a plant, e.g. vernalization response, flowering time, leaf shape and/or shade avoidance response, the method including causing or allowing anti-sense transcription from heterologous

nucleic acid according to the invention within cells of the plant.

The present invention further provides the use of the
5 nucleotide sequence of VRN2, or a fragment, mutant,
derivative, allele, variant or homologue thereof for down-
regulation of gene expression, particularly down-regulation of
expression of a VRN2 gene or homologue thereof, preferably in
order to influence a physical characteristic of a plant,
10 especially vernalization response, flowering time, leaf shape
and/or shade avoidance response.

When additional copies of the target gene are inserted in
sense, that is the same, orientation as the target gene, a
15 range of phenotypes is produced which includes individuals
where over-expression occurs and some where under-expression
of protein from the target gene occurs. When the inserted
gene is only part of the endogenous gene the number of
under-expressing individuals in the transgenic population
20 increases. The mechanism by which sense regulation occurs,
particularly down-regulation, is not well-understood.
However, this technique is also well-reported in scientific
and patent literature and is used routinely for gene control.
See, for example, van der Krol et al., (1990) *The Plant Cell*
25 2, 291-229; Napoli et al., (1990) *The Plant Cell* 2, 279-289;
Zhang et al, 1992 *The Plant Cell* 4, 1575-1588.

Again, fragments, mutants and so on may be used in similar
terms as described above for use in anti-sense regulation.
30

Thus, the present invention also provides a method of
influencing a characteristic of a plant, e.g. vernalization
response, flowering time, leaf shape and/or shade avoidance
response, the method including causing or allowing expression
35 from nucleic acid according to the invention within cells of
the plant. This may be used to suppress activity of a product
with ability to influence vernalization response, flowering

time, leaf shape and/or shade avoidance response. Here the activity of the product is preferably suppressed as a result of under-expression within the plant cells.

5 Aspects and embodiments of the present invention will now be illustrated, by way of example, with reference to the accompanying figures. Further aspects and embodiments will be apparent to those skilled in the art. All documents mentioned in this text are incorporated herein by reference.

10

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows total leaf number (Rosette plus Cauline) of *Ler* plants (squares), *fca-1* plants (diamonds) and *vrn2-1 fca-1* 15 plants (circles) after various periods of vernalization (measured in days).

Figure 2 illustrates the Red Far-Red Light plant phenotype:

Figure 2A shows the area of the largest leaf of *Ler*, *fca-1* and *vrn2-1 fca-1* plants grown under white light (W) (open bars) or white light with supplementary FR light (W + FR) (shaded bars).

Figure 2B shows the rosette leaf number at bolting of plants treated as in the experiment of which results are shown 25 in Figure 2A.

Figure 3 illustrates results of genetic and physical mapping of VRN2 in *Arabidopsis thaliana*.

30 Figure 4 illustrates cosmids and genes near VRN2 in the *Arabidopsis* genome.

Figure 5 illustrates the structure of the VRN2 gene of *Arabidopsis*, the 5K cDNA aberrant splicing, and the position 35 and nature of the *vrn2-1* mutation. Exons are shown as open boxes. Untranslated regions are shown as shaded boxes. Introns are shown as lines.

Figure 6 shows the sequence of VRN2 cDNA, including coding sequence and predicted amino acid sequence of the encoded protein. Putative NLSs are boxed, the putative acidic activation domain is underlined. The putative zinc-finger motif is doubly underlined. The positions of introns are indicated with arrows. The position of the *vrn2-1* mutation is circled.

Figure 7 illustrates dCAPS Marker for the *vrn2-1* Mutation. A diagnostic derived CAPS (dCAPS) marker was designated for the *vrn2-1* mutation. This utilizes a primer (VRN2-AZ) that includes half of the recognition site for the XmnI restriction enzyme, the other half is supplied, specifically, by the sequence of the *vrn2-1* mutation. This results in a successful restriction digestion only when using PCR amplified genomic DNA from *vrn2-1* mutants as a template.

Figure 8 shows alignment of the *Arabidopsis* VRN2 amino acid sequence with similar proteins.

Figure 8A aligns the full-length VRN2 protein with four other proteins, using Clustal method with PAM250 residue weight table, performed on 17 January 1999 at 19:19 GMT.

Figure 8B aligns the zinc finger region, using Clustal method with PAM250 residue weight table, performed on 17 January 1999 at 19:25 GMT.

Abbreviations:

At *Arabidopsis thaliana*, Sc *Saccharomyces cerevisiae*, Sp *Schizosaccharomyces pombe*, Ce *Caenorhabditis elegans*, Dm *Drosophila melanogaster*, Hs *Homo sapiens*, Mm *Mus musculus*, Rn *Rattus norvegicus*, Xm *Xiphophorus maculatus*.

LIST OF SEQUENCES

SEQ ID NO:\SEQUENCE

5 1 Landsberg erecta VRN2 cDNA
2 Landsberg erecta VRN2 amino acid
3 Landsberg erecta VRN2 genomic
4 Columbia VRN2 cDNA
5 Columbia VRN2 amino acid
10 6 Columbia VRN2 genomic
7 5K (Columbia, aberrant splice) cDNA
8 5K (Columbia, aberrant splice) amino acid
9 C72616 EST (modified) cDNA
10 C72616 EST (modified) amino acid
15 11 AI163743 EST (modified) cDNA
12 AI163743 EST (modified) amino acid
13 At Hyp 2245035 (ATFCA7_4) (modified) cDNA
14 At Hyp 2245035 (ATFCA7_4) (modified) amino acid
15 KIAA0160 cDNA
20 16 KIAA0160 amino acid

Additional sequences included in the Figures:

17 Landsberg erecta VRN2 zinc finger amino acid
25 18 At Di19 S51478 zinc finger 1 amino acid
19 At Di19 S54178 zinc finger 2 amino acid
20 At SUP U38946 zinc finger amino acid
21 At Hyp 2191171 zinc finger amino acid
22 At Hyp 3377806 zinc finger amino acid
30 23 Sc Pep7 91500 zinc finger amino acid
24 Sc TFIIB 730931 zinc finger amino acid
25 Sp Hyp 1351713 zinc finger amino acid
26 Ce Hyp 255942 zinc finger amino acid
27 Ce Hyp 2854197 zinc finger amino acid
35 28 Ce Hyp 304459 zinc finger amino acid
29 Dm BRCORE-NS-Z3 zinc finger amino acid
30 Dm GAGA 729556 zinc finger amino acid
31 Dm ken 3550814 zinc finger amino acid

32 Hs ATBF-1 976347 zinc finger amino acid
33 Hs KIAA0160 zinc finger amino acid
34 Hs ZNF142 3123312 zinc finger amino acid
35 Mm FOG 2252814 zinc finger amino acid
5 36 Mm Spalt 1296845 zinc finger amino acid
37 Rn Roaz 2149792 zinc finger amino acid
38 Xm ZF1 532083 zinc finger amino acid

EXAMPLE 1

10 *Characterisation and Cloning of VRN of Arabidopsis Thaliana and Mutant Alleles Thereof*

Isolation of vrn2 mutants

15 Two vrn2 mutant alleles (vrn2-1 and vrn2-2) were isolated by mutagenising fca-1 seeds with EMS as described by Chandler et al. (Plant J (1996) 10: 637-644). WO96/38560 (PCT/GB96/01332) discloses the sequence of fca and mutant alleles and their cloning and characterisation. The vrn2-1 fca-1 line used here
20 has been backcrossed to fca-1 four times. For mapping purposes, the vrn2-1 allele was used (at the 2nd backcross).

Phenotypic Characterization

25 *Vernalization*

The vernalization response of vrn2 mutant plants was investigated by examining their flowering time in response to increasing durations of vernalization treatment.

30 Standard vernalization conditions were used, i.e. low light intensity $5 \mu\text{mol m}^{-2} \text{ s}^{-1}$, 8 hr photoperiod, 5±1 degree C, for varying periods (from between 1 and 42 days). Similar effects would be observed under continuous or no light, the
35 temperature is more important.

In the absence of a vernalization treatment, *vrn2-1 fca-1* mutant plants showed a small but consistent delay in flowering compared to the parental (wild-type) *fca-1* controls (Figure 1). *vrn2-1* has a higher leaf number than *fca-1*, a reduced leaf 5 number correlating with a transition from vegetative to reproductive (i.e. flowering) state). However, following vernalization, this difference was greatly magnified (Figure 1). The response exhibited by *vrn2-1 fca-1* plants was typically a 35% reduction in total leaf number after 6 weeks 10 of vernalization, compared to 67% reduction of *fca-1* controls. The delay in flowering, both with and without a vernalization treatment, as measured by increased leaf number, was also observed if the days to flower i.e. the day at which the first floral bud was visible was used.

15

Red/Far-Red light perception

The response to vernalization in *Arabidopsis* has been positively correlated with the response to different ratios of 20 Red(R) to Far-red(FR) light; mutants and ecotypes that respond strongly tend to respond strongly to conditions of low R:FR (Bagnall, Ann Bot (1993) 71: 75-83; Martinez-Zapater et al., Plant Physiol (1990) 92: 770-776). This response typically manifests itself in two distinct ways - an acceleration of 25 flowering time (leading to an effect that mimics the effect of a vernalization treatment) and a reduction in leaf area (or shade avoidance). This response is believed to have evolved to allow plants to adapt to the availability of light allowing individuals to seek light when in competition with 30 their neighbors.

We examined the ability of *vrn2-1 fca-1* mutants to respond to conditions that mimic such an environment.

35 Under these conditions, *vrn2-1 fca-1* plants showed a marked reduction in the shade avoidance response, with the mean area of the largest leaf decreasing by only 26%, compared to 74%

for the *fca-1* control (Figure 2A). However, the *vrn2-1* mutation does not appear to affect all aspects of the response to FR light, as *vrn2-1 fca-1* plants showed a similar acceleration of flowering in response to supplementary FR 5 light as *fca-1* controls (Figure 2B).

These data provide indication that VRN2 plays a role in regulating the response to FR light, and may mediate changes specifically in leaf size (as flowering time is only slightly 10 affected) under conditions of low R:FR ratios.

Genetic Mapping

The VRN2 gene was mapped in an F2 population derived from the 15 cross *vrn2-1 fca-1* crossed to *fca-10*, following the procedure used to map the VRN1 gene (Chandler et al. *supra*). The VRN2 gene was initially positioned using a population of 70 F2 individuals between the RFLP markers g13683 and mi112 (Schmidt et al., Plant J (1996) 9: 755-765) on the long arm of 20 Chromosome IV (or D), using conventional techniques. The map position of VRN2 was further refined by screening an additional 429 F2 plants with a SSLP derived from the marker g19247 (Schmidt et al., (1996) Plant J 9: 755-765) and the CAPS marker g4539 (Parker et al., Plant Cell (1998) 9: 1-17). 25 A total of 12 individual F2 plants that were recombinant between g19247 and g4539 were further analyzed with RFLP markers g13683 and CC36F6 (Bancroft et al., Weeds World 4ii:(1997)) and the CAPS marker C18 (Parket et al., *supra*), and with two additional CAPS markers (VRN2RS, VRN2CD) 30 generated using the published Columbia sequence (Bevan et al., Nature (1998) 391: 485-488) (accession numbers Z97341 and Z97342) as a template. The VRN2 gene was localized to a 245 kb region defined at the centromeric (north) end by an RFLP detected with the cosmid 35 CC36F6, and at the telomeric (south) end by the g4539 CAPS marker. This interval is defined by 3 recombinant individual

F2 plants; 1 recombinant between VRN2 and CC36F6, and 2 recombinants between g4539 and VRN2 (Figure 3).

Physical Mapping

5

The genetic interval defined by CC36F6 and g4539 is almost completely covered by the 3 BAC clones T1C7, I5D3 and T5O15. Cosmids derived from subclones of YAC EW16B10 in the binary vector 04541 (Bancroft *et al.*, *supra*.) (abbreviated as IB)

10 were positioned by end sequencing, and ordered relative to the published sequence of the Columbia ecotype in this region (Genbank accession numbers Z97341 and Z97342). We selected cosmids clones that extended from the complex RPP5 locus outwards to CC36F6 and g4539, reasoning that VRN2 was not 15 within the RPP5 locus, which is comprised of multiple repeats of RPP5-like genes in both Columbia and Landsberg ecotypes (Bevan *et al.*, *supra*.).

Additional Landsberg cosmids in the 04541 binary vector 20 covering the region not covered by the Columbia YAC subclone cosmids were identified by hybridization to the inserts from BACs T1C7 and T5O15, and aligned based on end sequencing, and compared to the published Columbia sequence (Bevan *et al.*, *supra*) and to the sequence of the Landsberg ecotype in this 25 region. An almost complete cosmid contig was generated over this region.

Simultaneously with the isolation of cosmids, ordered cosmids, beginning with those at the centromeric end of the contig, 30 were transformed into *vrn2-1 fca-1* plants by *Agrobacterium tumefaciens*-mediated vacuum infiltration (Bechtold *et al.*, C R Acad Sci Paris (1993) 316: 1194-1199). (Figure 3). The presence of the cosmid in each transgenic line (T1 plants) was confirmed by a cosmid-specific diagnostic PCR, comprising an 35 insert specific primer (corresponding to a portion of the Columbia genomic DNA) and a primer present in the cosmid vector.

Cosmid Complementation

Cosmids introduced into *vrn2-1 fca-1* plants were tested for their ability to complement the *vrn2* phenotype. T2 seeds, 5 from individual T1 plants segregating kanamycin resistance at a 3:1 ratio, were sown on soil and vernalized for two (in some experiments three) weeks. Plants were then transferred to greenhouse conditions, and after ten days pricked out into individual compartments of divided trays. Total leaf number 10 was determined, and cosmids were scored as complementing if the segregation ratio of early to late plants (when compared to *fca-1* and *vrn2-1 fca-1* controls) plants was approximately 3:1.

15 Two Columbia cosmids (4A23, 2 out of 2 T1s; 6N1, 1 out of 1 T1) clearly complemented the phenotype of *vrn2-1 fca-1* mutants, with the earliest plants flowering at approximately the same time as vernalized *fca-1* plants.

20 Sequence Analysis and ORF Prediction

The sequence in common to both IB4A23 and IB6N1 cosmids has previously been annotated as containing 2 complete predicted genes (ATDL4445W and ATDL4450W) and (presumably 25 non-functional) portions of two other genes - the 3' end of ATDL4440W and the 3' end of the RPP5-like gene, CHPR (ATDL4460W) (Bevan et al., *supra*.) Genbank accession number Z97342. In addition, a cognate cDNA (5K) not included in the annotation is present in this region, and appears to span two 30 of the predicted genes (ATDL4445W and ATDL4450W) (Bevan et al., *supra*). However, as two cosmids in the region (1 independent T1 line from each of cosmids IB4N5 and IB6C5) did not complement the mutant phenotype (Figure 4) this ruled out the predicted gene ATDL4445W. This left the unannotated cDNA 35 5K, and the predicted gene ATDL4450W as candidates for VRN2. However, the presence the cognate cDNA 5K from the Columbia ecotype that overlapped both ATDL4445W and ATDL4450W

necessitated a re-examination of the prediction for the ATDL4450W gene.

In order to define the structure of these genes, we used the
5 NetGene2 prediction program (Hebsgaard et al., Nucl Acids Res (1998) 24: 3439-3452), using "Arabidopsis" as the organism option (the only parameter that can be set manually). BLAST,
PSI-BLAST, PSORT and PROSITE programs were used to identify potential function domains and similarities (Altschul et al.,
10 Nucleic Acids Res (1997) 25: 3389-3402; Bairoch et al., Nucl Acids Res (1997) 25: 217-221; Nakai et al., Genomics (1992) 14: 897-911). Default parameters of TBLASTN, PSI-BLAST and BLASTP were used (Expect = 10, BLOSUM62 matrix, gap penalty = 11, penalty per gap 1, lambda ratio 0.85). The NCBI/GenBank
15 database was used. The PSORT algorithm (Nakai) was used, using the "plant" option as the source organism (the only parameter that can be manually changed). The Profile Scan program at PROSITE (Bairoch) was used to search for motifs in VRN2, with default parameters (there are no parameters a user
20 can select, the results being "hit" or "no hit").

This analysis yielded predictions for two genes, 5K, a putative nuclear localized protein that is highly post-transcriptionally spliced (15 exons), represented by the
25 Columbia cognate cDNA; and a modified prediction for 4450, with 6 putative membrane-spanning domains, represented at its 3' end by an Arabidopsis EST (accession number T22412).

Determination of the vrn2-1 Mutation and Identification of the
30 VRN2 gene

In an attempt to determine which gene (5K or 4450) is VRN2, PCR primers were designed to amplify products encompassing the entire predicted open reading frame of both genes.
35

Three independent RT-PCR reactions using total RNA prepared from fca-1, vrn2-1 fca-1 and vrn2-2 fca-1 14 day-old seedlings

grown on GM plates in continuous light were performed for each predicted gene with a high fidelity enzyme mix (Boehringer Mannheim, HiFi System). These PCR products were sequenced using both the primers used for PCR, and a series of internal 5 primers, using the BIGDYE kit (PE Applied Biosystems). The reactions were run on an ABI377 machine, and compiled using the SeqMan (DNASTar, Lasergene) program.

The sequences of the PCR confirmed our prediction for both 10 genes, and indicated that we had amplified across the entire open reading frame of 5K and ATDL4450W as anticipated.

Several minor polymorphic differences were detected between the published Columbia sequence, and the Landsberg erecta 15 sequence we amplified by PCR. These differences were consistent with the Landsberg erecta genomic sequence in this region. Furthermore, the Columbia cDNA for 5K appears to utilise a different splice donor site from that used in the Landsberg ecotype, and would produce a truncated, presumably 20 non-functional protein (Figure 5). However, we have also sequenced the Columbia 5K product derived independently by RT-PCR, and this appears to use the same splice site as Landsberg, and should encode a functional protein. A consistent difference between the *vrn2* mutants and *fca-1* was 25 detected in the 5K PCR product, a G to A change at position 1201 of the predicted cDNA in *vrn2-1 fca-1* (Figure 5). We are currently investigating the nature of the mutation in the *vrn2-2* allele. This type of mutation, a single base-pair 30 change, is commonly observed following EMS mutagenesis. This mutation converts a TGG codon (Tryptophan) to stop codon (TGA), and would result in the production of a truncated protein of 322 amino acids in the *vrn2-1* mutant, compared to 443 amino acids of wild type VRN2 (Figure 6). The presence 35 of this mutation indicated that 5K was likely to be VRN2. The presence of the *vrn2-1* mutation in the genome of *vrn2-1 fca-1* mutant plants was confirmed by a derived CAPS (dCAPS) (Michaels et al., Plant J (1998) 14: 381-385; Neff et al.,

Plant J (1998) 14: 387-392) marker specific for the *vrn2-1* mutation (Figure 7). This diagnostic test is specific for the *vrn2-1* mutation, as it detects wild type VRN2 in both *fca-1* and *vrn2-2 fca-1* mutants.

5

Analysis of the VRN2 gene

To gain an insight into the possible function of the VRN2 gene, and how the *vrn2-1* mutation may affect the function of 10 the VRN2 protein, we compared the amino acid sequence of VRN2 to several databases of protein and translated nucleic acid sequences using BLASTP and TBLASTN programs at NCBI, using the default parameters as noted above.

15 Several molecules with a significant degree of similarity were identified (Table 1 and Table 2).

One such gene, represented by a human cDNA (KIAA0160) shares homology with VRN2 over a short region near the amino terminus 20 of VRN2 (amino acids 63 to 132), and a longer, but less conserved region of homology towards the carboxy terminus (amino acids 263-366) (Figure 8a). Closer examination of the amino terminal conserved region revealed that it matches the consensus of a zinc-finger motif. Such motifs can take a 25 variety of forms, but all co-ordinate zinc atoms through two cysteine residues, and two cysteine or histidine residues. VRN2 falls into the latter class, having a C2H2 motif comprised of two cysteines separated by 2 amino acids, and two histidines separated by two amino acids (Mackay et al., TIBS 30 (1998) 23: 1-4) .

Zinc-finger motifs are known to be capable of mediating both protein-DNA and protein-protein interactions. The zinc finger motif of VRN2 does not closely resemble the large EPF family 35 of *Arabidopsis* C2H2 zinc fingers from *Arabidopsis*, which have a highly conserved QALGG motif in the middle of the zinc finger (Kubo et al., Nucl Acids Res (1998) 26: 608-615) (Figure

8b). In addition, VRN2 differs from the EPF proteins in that VRN2 has a single zinc finger motif, whereas most members of the EPF family (with the exception of SUP and AtZFP1) have between two and four zinc fingers (Kubo *et al.*, *supra*). This 5 amino-terminal region (amino acids 63-132), and particularly the zinc-finger motif (amino acids 90-111) may thus represent a domain that mediates protein-protein or protein-DNA interactions.

10 The carboxy terminus of VRN2 (amino acids 263 to 366) is similar to several other candidate genes (Table 1 and Table 2). As mentioned above, there is limited homology to the human predicted protein KIAA0160. The molecule showing greatest homology to VRN2 is an EST sequence from poplar 15 (*Populus tremula L. x Populus tremuloides* Michx (Accession Number AI163743) (Sterky *et al.*, PNAS USA (1998) 95: 13330-13335) which has 52.8% identity over 127 amino acids (Table 1), as calculated with the BLASTP algorithm using default parameters (as noted above). VRN2 also shows 20 significant similarity to a predicted *Arabidopsis* protein (ATFCA7_4, Accession Number 2245035) (Bevan *et al.*, *supra*), which is quite close to VRN2 on chromosome 4, only 30 kb away towards the centromere. A close examination of the sequence near this gene revealed that the prediction as annotated may 25 be incorrect, as the use of a different splice site, resulting a different carboxy terminus to the protein, increases the degree of homology with VRN2. The similarity of these two *Arabidopsis* genes raises the possibility that VRN2 may be a member of a gene family in *Arabidopsis*, and their close 30 position suggests that these genes may have arisen following a duplication. Arguing against this notion is the observation that these two genes, VRN2 and ATFCA7_4, are transcribed in opposite directions. A rice EST (C72616) also shares significant similarity with the carboxyl region of VRN2, 35 suggesting that this region may form an evolutionarily conserved domain present in monocots and dicots.

This conserved carboxy region is predicted to be highly charged, as it is composed of a large number of acidic residues (D and E). This highly charged region is highly similar in the Poplar and rice ESTs (Table 2). Such acidic regions are found in a number of eukaryotic transcription factors, and often function as activation domains (Hahn, Cell (1993) 72: 481-483). It is therefore possible that VRN2 may function as a transcription factor, given it has both a DNA-binding motif (or protein binding) (amino acids 63-132) and a putative activation domain (amino acids 263-328). Furthermore, the amino portion of VRN2 contains two predicted nuclear localization signals (NLSS) (Figure 6). The first is a simple 4 residue basic signal, while the second is a bi-partite signal, that fits the consensus (R/K)(R/K)N10(F/K)4 (Dingwall et al., TIBS (1991) 16: 478-481).

EXAMPLE 2

Production and Characterisation of Arabidopsis Transgenic for VRN2

VRN2 cDNA in the sense orientation is cloned into plant expression vectors SLJ4D4 and SLJ4K1 (Jones et al., (1992) Transg. Res. 1, pp 285-297) according to the teaching of Sambrook J et al (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. SLJ4D4 and SLJ4K1 place the VRN2 cDNA under the control of the Ca MV 35S promoter and include terminator sequences from octopine synthase (ocs) and nopaline synthase (nos) genes, respectively.

Antisense constructs are produced in the same manner except that the VRN2 cDNA is inserted into the expression vectors in the opposite orientation.

The VRN2 expression cassettes are then subcloned separately into the binary vector SLJ1714 (Jones et al., *supra*), and mobilised into *Agrobacterium* strains by tri-parental mating

according to the teaching of Hoekema et al., (1983) Nature 303, pp 179-180. *Arabidopsis* are transformed with the *Agrobacterium* strains carrying VRN2 expression constructs (either sense or antisense) following the teaching of Bechtold et al., (1993) C R Acad. Sci. Paris 316, pp 1194-1199.

Arabidopsis plants are assayed for changes in their response to changes in the ratio of far red light, essentially as described by Halliday et al., (1997) Plant J. 12, pp 1079-1090.

Results

Differences in vernalisation requirement and response are observed in *Arabidopsis* plants transgenic for VRN2 (sense or antisense orientation) relative to *Arabidopsis* plants transformed with empty vectors and non-transformed *Arabidopsis* plants.

20 EXAMPLE 3

Production and Characterisation of Tobacco Transgenic for VRN2

VRN2 cDNA in the sense orientation is cloned into plant expression vectors SLJ4D4 and SLJ4K1 as described in Example 25 2.

Antisense constructs are produced in the same manner except that the VRN2 cDNA is inserted into the expression vectors in the opposite orientation.

30

VRN2 expression cassettes subcloned separately into the binary vector SLJ1714 (Jones et al supra), are mobilised into *Agrobacterium* strains by tri-parental mating as in Example 2, and tobacco plants are transformed using the *Agrobacterium* strains carrying VRN2 expression constructs (either sense or

antisense) following the teaching of Horsch et al., (1985) Science 227, pp 1229-1231.

Tobacco plants are assayed for changes in their response to
5 changes in the ratio of far red light, essentially as described by Halliday et al., (1997) Plant J. 12, pp 1079-1090.

Results

10 Differences in vernalisation requirement and response are observed in Tobacco plants transgenic for VRN2 (sense or antisense orientation) relative to Tobacco plants transformed with empty vectors and non-transformed Tobacco plants.

15 **EXAMPLE 4**
Production and Characterisation of Brassica (Oil Seed Rape, Winter Type) Transgenic for VRN2

20 Sense and antisense constructs are used to generate Agrobacterium strains carrying VRN2 expression constructs (either sense or antisense) as described in Example 2 and Example 3, and are used to transform oil seed rape following the teaching of Moloney et al., (1989) Plant Cell Rep. 8, pp
25 238-242.

Results

Differences in vernalisation requirement and response are
30 observed in oil seed rape plants transgenic for VRN2 (sense or antisense orientation) relative to oil seed rape plants transformed with empty vectors and non-transformed oil seed rape plants.

EXAMPLE 5*Production and Characterisation of Rice Transgenic for VRN2*

VRN2 cDNA in the sense or antisense orientation is cloned into
5 constructs and used to generate respective *Agrobacterium* strains. The *Agrobacterium* strains carrying VRN2 expression constructs (either sense or antisense) are used to transform rice following the teaching of Kohll A et al (1998) Proc. Natl. Acad. Sci. USA 95, pp 7203-7208.

10

Results

Differences in vernalisation requirement and response are observed in rice plants transgenic for VRN2 (sense or
15 antisense orientation) relative to rice plants transformed with empty vectors and non-transformed rice plants.

EXAMPLE 6*Production and Characterisation of Wheat Transgenic for VRN2*

20

Agrobacterium strains carrying VRN2 expression constructs (either sense or antisense) are generated as in preceding Examples and are used to transform wheat following the teaching of Becker D et al., (1994) Plant J. 5, pp 299-307.

25

Results

Differences in vernalisation requirement and response are observed in wheat plants transgenic for VRN2 (sense or
30 antisense orientation) relative to wheat plants transformed with empty vectors and non-transformed wheat plants.

METHODS AND MATERIALS35 **Plant Growth**

For vernalization treatments, seeds were sown on a damp layer

of fine grit (Levington's M3) on wet soil in individual pots, and vernalized for increasing durations at 4°C, 8hr light:16hr dark, 5 $\mu\text{mol m}^{-2} \text{ sec}^{-1}$ light intensity. Seed sowing was staggered, with all plants removed from the vernalization conditions simultaneously. Following vernalization, plants were placed into a controlled environment chamber (Gallenkamp), 20°C, 16 hr light: 8hr dark 90 $\mu\text{mol m}^{-2} \text{ sec}^{-1}$ light intensity. Plants receiving no vernalization treatment were stratified for 2 days under vernalization conditions, and grown for two days prior to transfer into the growth cabinet. Plants were grown for 10 days, and then pricked out into individual compartments of P40 trays. The flowering time, as measured by counting total leaf number (i.e. rosette and cauline leaves) was determined once the primary inflorescence had elongated sufficiently.

The phenotype of *vrn2-1 fca-1* plants was examined under different ratios of red to far-red (R:FR) light. Plants were stratified for 2 days, then grown for 10 days under continuous light, and then transferred to separate growth chambers with white light (W), R:FR ratio of 5.8, with or without supplementary FR light (W+FR), R:FR ratio 0.08. The number of rosette leaves was determined to measure flowering time, and the area of the largest leaf was used as an indicator of the shade-avoidance response.

Mapping

VRN2 was positioned on the long arm of Chromosome 4 (or D) through linkage to the RFLP marker m506 (Chang et al., Proc Natl Acad Sci USA (1988) 85: 6856-6860), in progeny of a cross between *vrn2-1 fca-1* (*Ler* background) and *fca-10* (*Ws* background). Further RFLP markers in this region (Liu et al., Plant J (1996) 10:733-736) were used to refine the position of *VRN2*. We used standard RFLP techniques, using ^{32}P -labeled cosmid probes, and a PhosphorImager detection system. A *Ler:Ws* RFLP was detected when using EcoRI digested genomic DNA, and either g13683 or CC36F6 as probes. The g19247 marker is a SSLP marker (using the primers g19247F and g19247R, see

below), with Ler producing a PCR-amplified band of approximately 750 bp, while Ws produces a band of 862 bp. Fine mapping of VRN2 was performed using a series of PCR-derived markers, based on the Columbia genomic sequence in 5 this region (Bevan et al., *supra*). The VRN2CD marker was amplified with two primers, VRN2-C and VRN2-D, and digested with DdeI. This produces a CAPS marker, with Ler producing bands of (approximately) 480 bp, 290 bp and 190 bp, while Ws produces bands of (approximately) 330 bp, 290 bp and 150 bp. 10 The VRN2RS marker is a dominant marker for Ws (i.e. Ler:Ws heterozygotes cannot be distinguished from Ws), produced by amplifying genomic DNA with VRN2-R and VRN2-S (see below), and digesting the product with *Mbo*II. This produces a single predominant band (and several much smaller unresolved bands) 15 in Ler of approximately 400 bp, while Ws produces two bands, 400 bp and 300 bp.

Cosmid Isolation

Cosmids covering the region not covered by the subclones of 20 EW16B10 YAC were identified through hybridization to BAC inserts derived from BACs T5015 and T1C7. BAC DNA was purified and the insert isolated following digestion with *Not*I and separation by pulsed field gel electrophoresis (PFGE) as described in (Bancroft et al., *supra*). Purified BAC inserts 25 were labeled with a³²P-dCTP by random primed labeling, and hybridized to arrayed grids of a Ler genomic cosmid library. Positively hybridizing cosmids were identified, and rescreened following restriction enzyme digestion, southern blotting and hybridized to the BAC insert probe used initially. Cosmids in 30 the region of interest were selected, and genomic sequence was obtained from the ends of the insert using the BIGDYE cycle sequencing kit, and T3 and T7 primers, whose sequences flank the genomic DNA insert site. This sequence was aligned with that of the Columbia genomic sequence to accurately position 35 the cosmids.

Complementation

Cosmids in the 04541 binary vector were mobilized into *Agrobacterium tumefaciens* (strain C58C1 RifR by tri-parental mating (Hoekema et al., *Nature* (1983) 303: 179-180). *vrn2-1*

5 *fca-1* plants were transformed with these *Agrobacterium* strains by vacuum infiltration (Bechtold et al, *supra*.). Transgenic T1 plants were selected on GM with Kanamycin (50 mg/mL), and transferred to soil when they had reached the 3-4 leaf stage. The presence of each cosmid in the transgenic lines was
10 confirmed using either a Ler/Col specific polymorphism (CAPS or SSLP marker) or more commonly, through the use of a specific diagnostic PCR reaction, using a primer present within the cosmid insert sequence and a primer present in the cosmid flanking the insert site. Transgenic plants were also
15 tested for the presence of the *fca-1* mutation , as the T0 plants should carry this mutation. Together, these two methods ensured that only *vrn2-1 fca-1* plants carrying the desired cosmids were analyzed further. We aimed to produced 5 to 6 independent T1 transformants for each cosmid, but for
20 some cosmids only a single line was produced. However, as complementation was observed before we had generated T1s for all of the cosmids, we are continuing to produce additional transgenic T1 plants carrying the cosmids near VRN2. T1 plants were grown in a controlled environment chamber
25 (conditions as above), and allowed to self. T2 seed were collected, and analyzed for the segregation of Kanamycin resistance or sensitivity on GM plates containing Kanamycin (as above), scored 14-20 days after germination. Progeny from T1 plants that segregated a 3:1 ratio of resistant to
30 sensitive plants were tested for their ability to complement the *vrn2-1* mutant phenotype, by vernalizing for 3 weeks and recording the total leaf number.

Sequencing ORFs

35 Two potential open reading frames were sequenced following RT-PCR using RNA isolated from *fca-1* and *vrn2-1 fca-1* plants. cDNAs were reverse transcribed, primed by a dT12-18 primer,

and specific primers were then used to PCR amplify regions corresponding to the ORFs for ATDL4450W and 5K (VRN2). The Boehringer Mannheim HiFi PCR system was used to increase the fidelity of amplification. The primers VRN2-AL and VRN2-AM
5 were used to prime ATDL4450W, and VRN2-AI and VRN2-AJ for 5K. For 5K, the PCR reaction for the full length product (VRN2AI-VRN2AJ) was inefficient, so subsequent reactions were performed to amplify the cDNA in two overlapping fragment, using the primer combinations VRN2-AI with VRN2-AS; and
10 VRN2-AO with VRN2-AJ. The PCR products were isolated and purified, and sequenced directly using the BIGDYE sequencing kit (PE Applied Biosystems). At least two independent PCR products were sequenced from each allele. We sequenced the ATDL4450W PCR products with the amplifying primers VRN2-AL
15 VRN2-AM, and VRN2-AU through VRN2-AX. The 5K cDNA was sequenced with VRN2-AI, VRN2-AJ and VRN2-AO through VRN2-AT. Sequences were aligned into contigs using the DNASTar software package (LaserGene).

20 Sequence Comparisons

EST and cDNA sequences were first translated using the MapDraw program of DNASTar (LaserGene). This revealed several regions where the sequence appeared to be incorrect, as small changes to the described sequences drastically improved the similarity
25 to VRN2. Modified nucleic and amino acid sequences were initially aligned using programs within the MegAlign Package of DNASTar (LaserGene), as follows:

Amino acid sequences were initially aligned using the Clustal
30 V method (Higgins and Sharp (1989) CABIOS vol. 5, no.2, 151-153). Default parameters were used: Gap penalty 10, Gap length penalty 10, ktuple 2, and the PAM 250 residue weight table. Alignment was refined by eye and hand, making minor adjustments to position of gaps to improve alignment.

35

Nucleic acid (EST and cDNA) sequences were initially aligned with the Hein method (Hein (1990) Methods in Enzymology, vol. 183, 626-645), using default parameters: Gap penalty 11, Gap

length penalty 3, ktuple 2, and the weighted residue weight table. Where the nucleotide alignment (when translated) did not correspond to the amino acid alignment, the positions of gaps were adjusted to correspond with the gaps in the amino acid alignments.

dCAPS Marker for the vrn2-1 Mutation

A derived CAPS (dCAPS) marker was designed that was specific for the vrn2-1 mutation. Following PCR amplification from 10 genomic DNA with the primers VRN2-AY and VRN2-AZ, the 170 bp product was digested overnight with XmnI restriction enzyme, and the products resolved on a 4% agarose gel. Wild type plants (VRN2) produce a single band of 170 bp following digestion, while vrn2-1 mutants produce two bands, of 137 bp 15 and 33 bp.

Primers Used to Identify VRN2

As indicated on Figure 3 and Figure 5. All primer sequences are indicated 5' to 3'.

20 g19247F ACT GTT CGT CTC CTT CAT CAT G
g19247R TTG CTT GCC TGA AAA AAG TAT G
VRN2-C TGT CGA TAT GCG ACC AGT ACC
VRN2-D CAG GCT TAG ACC CAA TTG ACC
25 VRN2-R AGG TAG GAT CCG ACA TCG TCT TAT TTA CCG
VRN2-S CTC TTG AAT TCA AAA CTA TTC CTA CTC TCA CAC
VRN2-AI GCC AAT CGG TGT TTT CGC AGC TTT C
VRN2-AJ AAG AAT AAG TTA CAA TCC GAT AAA TCG G
VRN2-AL CAG TGG TTG AAG CTT AAG GAG G
30 VRN2-AM GCA ATG AAT AAA TCA TAA TCT TGG
VRN2-AO TCT ACT GGG ATG GTA GTT TTC
VRN2-AP ATA TCC CGA GGC AAC AGA GCT TG
VRN2-AQ CAT CTT TGG AAC TCG TTT G
VRN2-AR CTC AGT TGT AAT AGT TGC CC
35 VRN2-AS AAG AGT GGG CTA TGG CTG G
VRN2-AT GCA ACT CTT TCT CGT AAA ATC TTG
VRN2-AU GCC TCC ATA ACT GTC ATC ACA TC
VRN2-AV TTT CAT TGG TCA TGG GAT GG

VRN2-AW GAC TTC AGA GAT GGG TTT ATG C
VRN2-AX TCC ATA TCT AGC TCC TTC GCC
VRN2-AY TGC GTT CAT TAA GTA GGC AAC AGA AAA TGG
VRN2-AZ GAG AAG TAG TTA CCT TTG TTT TCT TAC AGA AGA GT

Table 1 Comparison of VRN2 Nucleotide and Amino Acid Sequence to Database Sequences

Sequence	VRN2 Col	Nucleotides				Amino Acids		
		Identity (%)	Length	Range ^a	Identity (%)	Similarity (%) ^b	Length	Range ^c
C72616	Region I	96.5	1722	1-1722	96.1	96.8	445	1-445
C72616	Region II	32.4	219	681-899	15.1	42.5	73	151-223
C72616	Complete	71.7	247	951-1197	72.7	85.7	82	241-322
A1163743	Region I	47.7	517	681-1197	40.1	58.1	172	151-322
A1163743	Region II	41.0	61	839-899	27.3	68.1	22	202-223
A1163743	Complete	71.6	264	951-1214	66.7	77.1	88	241-328
A1163743	Complete	65.8	376	839-1214	52.8	66.6	127	202-328
At Hyp 2245035		66.3	570	924-1493	63.7	79.5	190	232-421
KIAA0160	Region I	36.6	396	231-626	20.4	39.4	132	1-132
KIAA0160	Region II	35.5	904	819-1722	17.7	43.4	249	197-445
KIAA0160	Complete	41.1	1492	231-1722	16.0	36.0	445	1-445

^a Numbered relative to VRN2 cDNA sequence^b Similarity defined as identity plus similarity on the basis of amino acids grouped into four classes - (D, E), (R, K, H), (S, T, N, Q, Y), and (L, I, V, M, A, G, W, F, P, C)^c Numbered relative to VRN2 amino acid sequence

Table 2 Putative Domains of the VRN2 Protein

DNA/Protein-Binding	"Sub-domain"	Length	Position ^a	Identity (%)	Similarity (%) ^b	Molecule
Zinc Finger Motif	22	90-111	50.0	68.1	Mm Spalt 1296845	KIAA0160
	22	90-111	45.4	68.1	Sc TFIIB 730931	
	22	90-111	40.1	72.7	Ce Hyp 2854197	
	22	90-111	36.4	50.0	KIAA0160	
Conserved Region	41	76-116	43.9	61.0	KIAA0160	
Activation	66	263-328	89.4	95.4	AJ163743	
	68	263-330	83.0	93.2	C72616	
	104	263-366	58.6	76.0	At Hyp 2245035	
	104	263-366	29.8	49.0	KIAA0160	
Conserved Acidic Region	36	281-316	94.4	97.2	AJ163743	
	36	281-316	86.1	94.4	C72616	
	36	281-316	58.3	72.2	At Hyp 2254035	
	36	281-316	30.5	50.0	KIAA0160	

^a Position relative to VRN2 Amino Acid sequence
^b Similarity was calculated as defined in Table 1

SEQ ID NO: 1

CAAGCTTCTTCAATTGGCTCTCTTACACAGCCAATCGGTGTTTCGAGCTTCA
GGCCTCAATCCAAGACATTCTATATAAGCATATTGCAGAAGAGGCCGTTCTAATTGTTGCAT
5 TGAGTTATCGCTATGACGTAGGGAAATTCTAATTAGGGGAGGCCTCAGAGTTGCACCAA
CTTCATAATCGGCTTGTGACGTTGAGTGTAAATTGAACAAGAACAGAATTGTGAGGCAGAATTGT
CGCGCGAAATCCTCACCGGAGGAAGTGATTTCAACTGATGAGAACATCTTGATATATTGTAA
ACCTGTTGACTATATAACATCTTCACCTCGCTCTAGGCAACCCATCGTTCTCCAA
GATGCTTGAACACAAAATTGGAGCAAAGCGAAAAGAAAGTCAGAGATCTACTGGGATGGTA
10 GTTTCAACTATAAGGATTGTAAATAACACATTACAGAAAATGAAGTTAGGGAGGATTGTT
TTGTCATTTGCTCATGCTATGTGGTAGCTCAAGGGGCTGCAATTCAATTGAATT
CTCATGATTATTTGAATTGAGTTCAAGCTTCAAGAACAGACAGTTAACATGTTCT
GTAAAACCTTAATTCTCATATTGAGGAAGAAGGAAGTGATGACGATAAATTGAGCCCTT
CTCTCTGCTCGAAACCTCGTAAGCGGAGACAAAGAGGTGGCAGAAATAACACCAGGAGAC
15 TTAAAGTATGCTTTACCGTTGGATTACCCAGTTAACTAATGGCACAGAAAATGGAATC
ACCCTACTTAATGATGGAAACCGTGGTTAGGATATCCCAGGCCAACAGAGCTGCTGGACA
ATTGAGATGACCAGAACATTCCACCAGCCATAGCCACTCTCTGGACGCTGGTCTGA
AAGTTATATTGACAAGCGAAGCTGTGGTCCCTGCTACTAACAGAACAGAAAGTTATCTGCTGAG
CGATCAGAGGCTAGAAGGCCACCTACTTCTCAGAAACGCCATTCTATCATTCTCACAGAGT
20 CCAGCCAATGGCGCTTGAGCAAGTAATGCTGACCGGGATAGCGAGGATGAAGTCGATGACG
ATGTTGCAGATTTGAAGATGCCAGATGCTGACTTGTGGATGTGAATAAGATGAA
AAGCAATTATGCATCTTGGAACTCGTTGTAAGAAAACAAAGGGTTATAGCAGATGGTCA
TATCTCTGGGCATGTGAAGCATTTCAGAAAGATTTACGAGAAAGAGTTGCACCGTTACTCAT
CACTCTTCTGGTGGAGATTGTTTGATTAAACTATGGAACCATGGACTTGTGACTCA
25 GCCACCACATCAACAAACTGCAATACCATCCTCGAGAATTGCCGAATAGCTCAGACACCAC
CACCAACAAACAACACAGTGTGGATCGTCCAGTGAATCAACACCAACAATAACATTG
TGGATCATCCAATGACATAAACACAAGAACAAATGTTGACAACAAGGACAATAACAGCAGA
GACAAAGTAATTAAATAGGAAATCTCCGGCTTTATGATAACGATTATCGGATTGTA
TATTCTTCTTCTTAAAAATTGTTAGGAGCAAACAAATTGTTATGTTAGTGTATTCA
30 ACTGATTACATTTTAGTTAAAAAAATGGATTCTGCTTATAACT

SEQ ID NO: 2

35 MCRQNCRAKSSPEEVISTDENLLIYCKPVRLYNIFHLRSLGNPSFLPRCLNYKIGAKRKRS
RSTGMVVFNYKDCNNTLQKTEVREDCSCPFCMLCGSFKGLQFHLNSSHDLFEEFEKLFEEY
QTVNVSVKLNSFIFEEEGSDDDKEPFSLCSKPRKRRQRGGRNNTRLKVCFLPLDPSLTN
GTENGITLLNDGNRGLGYPEATELAGQFEMTSNIPPAIAHSSL DAGAKVILTSEAVVPATKT

RKLSAERSEARSHLLQKRQFYHSHRVQPMALEQVMSDRDSEDEVDDDADFEDRQMLDDFV
DVNKDEKQFMHLWNSFVRKQRVIADGHISWACEAFSRFYEKELHRYSSLFWCWRLFLIKLWN
HGLVDSATINNCNTILENCRNSSDTTTNNNSVDRPSDSNTNNNIVDHPNDINNKNNVDN
KDNNSRDKVIK

5

SEQ ID NO: 3

AAAGAGAACATGCTTGAACACTCTCATTGGCAAAACCTGACTGTATTATATCGCTTATTGTGT
10 GGTAAAGTTGACCTTGACTTTACAAGTTGGCGTTAAGAAGAGAGATGCGTAGATCAGCG
AGTGGTTGAGAGAGTTGGATCATTTCCCCGACTTCACGGTCTCCACGTCGATCTCAGAG
CATTACATCATTGAAGATGATGTGGAGGTGCTTGCCTAGGTACGATCCGAATTCTCAAG
CGGGGAAGAGAGAGAAGTCAAGATTGAGAAAACGTCATCCATTGATTCCCTCTC
ATTCTCTCTGTATCGCAATCCTCTGGCTCTCCTTATTCAAGGTAAGCCGAGAAATTG
15 ATTCAATCTCTATGAATCCATAATTGATATGTGAAACTTAATTAGGGATTTACAAAGGCTC
ATATGGATATGATATGAGGATCGAGATGTCCTGTAAACATTAGAATCTTGTGTTGAATTATT
GTTTCAATTGTTCATATTATACTAAACCGGTGATGGATTGGAATTGTCAGCAGCGTTAA
GGAGTTGAGTTCAAGAACATGTTGCTTGTCTCCATGGAACTCATCATATTCAAGTTT
TGGGAAAGGAAACAATTTTTACGCCGGTATTATGTGCCGCAAACCACATCGTAACATT
20 TGTAATTTCGGTTCTGTAGACACATAAAAGGATCTCTCGTTTACATGAAATGTATGTTAA
TATTTCACTATACATCACACAACTCAAGTAGAAAACACTGATGGTTATCCATTAAATCATCAT
TCTATTGGTCGAAAACAAGGATTAGTTCAACTTATTGCTACCTTAGTGATTAGATGTTCT
GTGAGTTTCAGCTAGCCAAGTCAACTAGAGTTAAACAATGGAATCAAATACATATTCA
ATTTATTAAACTCTGACTATTATGAAACAAAATGGAATTAAAATTGAAGGTACATGA
25 AGATTCTATTCTTAGTATGAAAAGTATAGATCAATGATAAAAGTATACCAACAGTGGT
GGATCTAGAAACATATTAGTATATGGCACAATATTTAACATATACAAATTAAATCTAA
AAGTTGTATTCAATTGAAAAGACWTCTGAATGAAGCAAATTATTGATGTGTTAAATCAT
CCATTATGTGTTAACGCCATTGATGTTAGTATAGTACTCTATGCTAACATAATT
ATACTATAAATTAAAAAAATAGGTAAGAAAAGAAAATAGATTAATATAAAAGCATT
30 TAGCTGAAATAAAATGAAAGAAGATAATAACTAATTGACTAAAAATTAGTAGAGCAT
ATGGGGCACAATACACTAAGTATTCATCTTACTATAAAATGTAACAAATTCAAAATTAT
CAAACGTATATAGGGCACGTGCCTAGGTACCAATAGACGTACGTCCGCCGTGAAATAAGTT
GGTGAATATGGTTAATTCTCTAAACTCACTGACTGCTACTGCCATGGTAGAGGTGAAAAAAC
AATTAGAAATTATAATGGATTAAGCTGTCCAAGTTGGTCGTATTTCTTACATT
35 TTAACATAAACATAAAAGTTCAACTATTGACTAGTAATAACGTGAAATGT
CTATTGGTTAAAATATGGCCATAAGGCCAGACTGAAAAAAACTGAAACCCAAAGT
TATATTCTACTGTTCTTCTCAGTGAATATCTCCAATCAAGCTCTCGATT
CTCTCTTACACAGCCAATCGGTGTTCGCAGCTTCAGGTTGTCTCAATCTCAAATTAA

AATCGGAGTCAAGTAATAACAATTGATAAACCTAATTGTTCCATTGTATTGTAAGATTGA
AATTTGCTGTAGATCCGAATCGAATTCTAGTTCTGGAATCGTGCATCGATGGAATT
TTTTTAAGATTCTCTTACACATTGGTCAAAAGATCACATAGTTTATTAAATTGA
TAAGTATGATGATTCTGCTAAGTGGCATTGGATAAAAGTTCGTTTGCAATACGTCTAA
5 CTTGTCTATGTCTTGAATGAACACTCTGAGTTGCTTAAAAAGTCTTGTGCTTCTTATTAC
ACAGGCCTCAATCCAAGACATTCTATATAAGCATATTGCAGAAGAGGCCGTTCTAATTGTTG
CATTGAGTTATCGCTATGACGTAGGGAAATTCTAATTAGGGAGGCCTCAGAGTTGCAC
TAACCTCATAATCGGCTTGTGACGTTGAGTGTATTGAACAAGAATGTGTAGGCAGAAT
TGTGCGCGAAATCCTCACCGGAGGAAGTGAATTCACTGATGAGAATCTCTGATATATTG
10 TAAACCTGTTGACTATATAACATCTTCACCTCGCTCTAGGCAACGTATGATTGCCT
TCCTCTCATCATATTAGCTAGTAATCTTCATCTCCTGTGAGTCACCCACTAATAGT
TTGAGTTGCTAAGCTGATTATGGTCTGATTGAGTGTGCTTCTTGTCTCCTA
AATTGAACTTGTGTTGTTGCAGCCATCGTTCTCCAAGATGCTTGAACTACAAAA
TTGGAGCAAAGCGAAAAGAAAGTATGTTCTTGAATGTAGCTGCTACAGTGATATGT
15 TATTTATCTTACTTCTAATATGAAAGCTGATGACCTATTTATCTTGTGAGTAGATATGG
ACATAATGAATGGTTCTTCTTGTGCTATGCTATAAACTTACATTAAATTGTTTT
GGTTAGGTCAAGATCTACTGGGATGGTAGTTCAACTATAAGGATTGTAATAACACATTAC
AGAAAACGTAAAGGTTAGTCTTCTGTCGACAAAATTGATGTCAATGTCTATGTT
CTCTAGATGATTGTTATTTACTATTTCTGATTGTCATGCAGTTAGGGAGGATTGTT
20 CTTGTCCATTTGCTCTATGCTATGGTAGCTCAAGGGCAACTATTACAACGTGAGGT
TTCTTCCGGGGCCTTCATATCTAACACTGTGAAATGCTACTGCTGTTCATGCTGTATACT
TTCACTGTTGGTACATATTTGTGTTGTTGTTCTCTCACTCTTCGAACGTGCT
GAGTGTGTGCTTATCTGAGAAAACATGTCCCAGATGGAGCTTACAACCAATTGCTTGTGTC
TATGCAGGGGCTGCAATTTCATTGAATTCTCATGATTATTGAATTGAGTTCAAGG
25 TATGTGGTTTATGAAATTCTGATTGCTATGCCTTATTAATGAGGTTATAGTTAAAAA
AGGGTCTTCCTATTGTAGCTTCGGAAGAACACCAGACAGTTAATGTTCTGAAACTTA
ATTCTTCATATTGAGGTCACTTAAACTGGTTATTGGAAATCCGATAGCTGGT
GAAAATTGTTATATTCCATCCTTATTGTACTAGGAAGAAGGAAAGTGTGACGATAAT
TTGAGCCCTCTCTGGTAACCTCAGAACCCCTCGATTAAACCTTAATAGCAGTA
30 ACTCCTGCTTCTCTGTCACTGCTAAATCCAACCTAAATGTTGAGCTCGAA
ACCTCGTAAGCGGAGACAAAGAGGTGGCAGAAATAACACCAGGAGACTAAAGTATGCTTT
TACCGTTGGATTCAACCGAGTTAACTAATGGCACAGAAAATGGAATCACCCTACTTAATGAT
GGTAAAATCATCTTCTGTGCGTTCTGTGGCTTAGAACTCATATTACAGAAGAAG
ATACAATGGCCTGATTGTTAGTTTGTACTTCTCCTCGATTCTTCTGCGAGGGTATTG
35 TTACCAAGAACTGATGTACAAATTAAATGGCATGCTACAGGAAACCGTGGTTAGGATATCCC
GAGGCAACAGARCTTGCTGGACAATTGAGATGACCAGCAACATTCCACCAACGCTG
CTCTTCTCTGGACGCTGGTCTAAAGTTATTGACAAGCGAAGCTGTGGTCCCTGCTACTA
AGACAAGAAAGTTATCTGCTGAGCGATCAGAGGCTAGAAGGTTGTTCATCATGACACCCCG

TCATCATAATTACCATTCCCTGTTACAATGTTCTTCTATTATGGATAAGTGTTATAG
TAATGCCATATTAACCGAGAAAATTCTTCCAGCCACCTACTTCTTCAGAAACGCCAATTCT
ATCATCTCACAGAGTCCAGGTGATCCAAGTTCTCACCTACTTCTTAGGCATTTCTTTA
AATTGCTCATGATGATATCTTATCAAAGCATACTGGTTGTTCTCATCCAAATTGTATTT
5 TGATCTGTATGTATCAACGAAAAATAGTTATGTCATGTTGTCCTCGTTTATTGCCACTAA
CCAAAAAAATGCATGTTCTGTGACAAGCCAATGGCGTTGAGCAAGTAATGTCTGACCGGGA
TAGCGAGGATGAAGTCGATGACGATGTTGAGATTGAAAGATGCCAGGTATTCCATGATT
TCTTCTGCGTTCATTAATAGACAACAGAAAATGGTATATGATGTAAC TGCTAATGGCTT
TTGAAACTAAAAAGCTGCAGATGCTGATGACTTGTGGATGTGAATAAAGATGAAAAGC
10 AATTCATGCATCTTGGAACTCGTTGTAAGAAAACAAAGGTAACTACTTCTCTAACACATG
AACAGACACAAAAAGACCTTATGTCTTACATTCCATACCTGTCTAAATGATTTGCTTATGG
AACTTGAGCTCAATTATGATTGTTGATGTTCAAGGGTTATAGCAGATGGTCATATTCTTGG
GGCATGTGAAGCATTTCAGGTTACGAGAAAGAGTTGCACCGTTACTCATCACTCTTCT
GGTAATATAAGTACACCAAACATATACAGACACATAACTACACTATCAATCTGTTCTGTT
15 TCTGAAAAAAAATTTCCAGGTGTTGGAGATTGTTTGATTAAACTATGGAACCA
TGGACTTGTGACTCAGCCACCATCAACAACTGCAATACCACCTCGAGAATTGCCGTAATA
GCTCAGACACCACCACCAACAAACAACAGTGTGGATCGTCCAGTGA CAAACACC
AACAAACAATAACATTGTGGRTCATCCCAATGACATAAACACAAGAACAAATGTTGACAACAA
GGACAATAACAGCAGAGACAAAGTAATTAAATAGGAAAATCTCCGGCTTTATGATACCGAT
20 TTATCGGATTGTAACTTATTCTTCTTCTTAAAAAATTGTTAGGAGCAAACAAATTTTTA
TATGTTAGTGTATTCAACTGATTACATTAGTTAAAAAAATGGATTCTGCTTATAA
CTAAAAACTGAAAAAAAGAAAAGTTCTTAATTCTTCTTGTACTGAGAAAAGCTCC
TCTAGTAAATATGAGTTATATTAAATCAAGTACATAACATAAAATAGTATATATTAAAGTG
CAAATAGATTGAAAACAAATCAAGAAGAAATTAAAGACAGAGTGATTAAGCTTAAACC
25 CCATTGGACTTGTCTTCTCAATGAATCCCTCACAAGCAGCAAGCTTCTCGATTTGCT
TTGACACCACCAATCGGTGTTTCGAATCTTCAGGTTGTCGATTCAATCTAGATCGG
AGTCAAGTAATAAAATTGATTAACCTAACAGTATTCCGTTCTCGTAAGAGTTGGGATTTAG
CACTAGATCGGAAATCGGAATTACGTTTGTAAAAGATTGATGGTTAGGTAATGGAAC
ATAGTTCTGGATTCAATTGCTTCTAGTTGATTCTCGAATTGTTGATTCGAATGCACATT
30 TTGTTCAAGGATCACAGAATTGATTAAAATTGACAAATTCCATCAATTCTCATAT
TAGGGTTTATATTCTTAGTAACTCGAACCTGTTGGAACTCTGTATACTCTGTGCTATGT
AGATAAAAGTCTAACATTGGTCAACTTGTTGATCTCTAAACTAGTTGGGCTCTCTGT
TTTAAAGTTGTGCTTCACTATTACACAGGTCTCATACAAGACTACAGTCTCAAGAAGCA
TAATATCGTCACTCTGTTGAGTTCTAACAGTGGTTGAAGCTTAAGGAGGTTCTTATG
35 TGCCTTTGATATC

SEQ ID NO: 4

CAAGCTTCTTCAATTGCTCTCTCTTACACGGCCAATCGGTGTTCGCAGCTT
CAGGCCTCAATACAAGACATTCTATATAAGCATATTGAGAAGAGGCGGTTCTAATTGTTGC
5 ATGGAGTTGAACAATATGACGTAGGGAAATTCTAATTAGGGGAGGCCTCAGAGTTGCAC
AACTTCATAATCAGCTCTGGACGTTGATTGATTGAACAAGAACAGTAGGAGAATT
GTCGCGCGAAATCCTCACCGGAGGAAGTGATTCAACTGATGAGAACATCTTGATATATTGT
AAACCTGTTGACTATATAACATCTTCACCTCGCTCTAGGCAACCCATCGTTCTGCC
AAGATGCTGAACATACAAAATTGGGGCAAAGCGAAAAGAACAGTCAAGATCTACTGGGATGG
10 TAGTTTCAACTATAAGGATTGTAATAACATTACAAAGAACTGAAGTTAGGGAGGATTGT
TCTTGTCCTTTGCTCTATGCTATGCTATGTTAGCTCAAGGGCTGCAATTGATTCATTGAATT
ATCTCATGATTATTGAATTGAGTTCAAGCTTTGGAAGAACACCAGACAGTTAATGTT
CTGTAAAACCTAATTCTTCATATTGAGGAAGAACAGGAGTGATGATGATAAATTGAGGCC
TTCTCTCTGCTCGAACCTCGTAAGCGTAGACAAAGAGGTGGCAGAAATAACACCAAGGAG
15 ACTTAAAGTATGCTTTACCGTTGGATTCAACCCAGTTAGCTAATGGCACAGAAAATGGAA
TTGCCCTGCTGAATGATGGAAACCGTGGTTAGGATATCCCAGGCAACAGAGCTTGCTGGA
CAATTGAGATGACTAGCAACATTCCACCAGCCATAGCCACTCTCTGGACGCTGGTGC
TAAAGTTATATTAAACAACCGAACGCTGTGGTCCCTGCTACTAACAGAACAGAAAGTTATCTGCTG
AGCGATCAGAGGCTAGAACGCCACCTACTTCTTCAGAAACGCCATTCTATCATTCTCACAGA
20 GTCCAGCCAATGGCGCTTGAGCAAGTAATGCTGATCGGGATAGCGAGGATGAAGTCGATGA
CGATGTTGCAGATTGAAAGATGCCAGATGCTGATGACTTGTGGATGTGAATAAGATG
AAAAGCAATTGACATCTTGAACTCGTTGTAAGAAAACAAAGGGTTATAGCAGATGGT
CATATCTCTGGGCATGTGAAGTATTTCAAGATTACAGAGAACAGTTGCACGTGTTACTC
ATCACTCTCTGGTGTGGAGATTGTTGATTAAACTATGGAACCCTGGACTTGTGACT
25 CAGCCACCATCAACAACGTAATACCATCCTCGAGAATTGCCGTAATACCTCAGTCACAAAC
AACACAACAAACAGTGTGGATCATCCCAGTGACTCAAACACCAACAATAACATTGTGGA
TCATCCGAATGACATAAAAACAAGAACAAATGTTGACAACAAGGACAATAACAGCAGAGACA
AGTAATTAAATAGGAAACACTCCGGTTAGATGATACCGATCTACGGATTGTAACATTATTC
TTCTTCTTAAAAAAATTGTTAGGAGCAAACAAAGATTATTGTTAGTGTATTCAACTG
30 ATTACATTAGTTAAAAAATGGATTCTCCTTAATAACT

SEQ ID NO: 5

35 MCRQNCRAKSSPEEVISTDENLLIYCKPVRLYNIFHLRSLGNPSFLPRCLNYKIGAKRKRKS
RSTGMVFNYKDCNNTLQRTEVREDCSCPFCMLCGSFKGLQFHNLNSHDLFEFEFKLLEEY
QTVNVSVKLNSFIFEEEGSDDDKFEPFSLCSKPRKRRQGRGRNNTRRLKVCFLPLDSPSLAN
GTENGIAALLNDGNRGLGYPEATELAGQFEMTSNIPPAIAHSSLDAAGAKVILTEAVVPATKT

RKLSAERSEARSHLLQKRQFYHSHRVQPMALEQVMSDRDSEDEVDDDVA
DVNKDEKQFMHLWNSFVRKQRVIADGHISWACEVFSRFYEKELHCYSSLFWC
HGLVDSATINNCNTILENCRNTSVTNNNNSDHPDSNTNNNNIVDHPNDIKNK
NNSRDK

5

SEQ ID NO: 6

AAAGAGAACAGAGCTTGACTCTCTCATTGGTCAAACCTGACTGTATTTATGC
10 GGTAAAGTTCGACCTTGACTTGACAAGTTGCCGTTAAGAAGAGAGATGCGTAGATCAGC
GAGTGGTTCTAGAGTTGGATCATTTCCGGCGACTTCAGGTCTCCGCCTCGATCTCAGA
GTGTTACAGCAATGGAAGAGATGATGTGGAGCTGCTTGCTAGGTACGATCCGAATTCTCAA
GCGGGAAAGAGAGAGAAATCAAGATTCAAGATTTCAGAAAACGTACCCATTGATTCCCT
CATTCTCTCTCTGTGTCGCAATCCTCTGGCTCTCCTACTCAGGTAAGCCGAGAAATT
15 GTTTCAATCTCTATGAATCCATAATTGATCTGTGAAACTTAATTAGGGATTACAAAGACT
CATATGGATATGAGGATCGAGATGTCTCTGCAACGTTAGAATCTTGTGTTGAATTATGGTTT
CAATTGTTCATATAACTAAATCGGTGATGGATTGGAATTGTCAGCAGCGTTAAGGAG
TTGAGTTCCAAAAGCAACATGTTGCTTGTCTCCATGGAACTCATATTCA
GGAAACAATTCTTTACCGCCGGTATTGTCAGCAGCGTTAAGGAG
20 GTTCTGTAGACACACAAAAGGATCTCGTTTCATGAAATGTATGTTAATATTCAGTGA
TATACATCACACAACCTCAAGTAGAAAACACTGATGGTTATCCATTAAATCATTCTATTGGTCG
AAAAAAAGATTAGTTCAACTTAATGCCACCTTAGGATTATGTTCTGTGAGTTCA
AGCCAACACTCAACTAGAGTTAACAAATGGAATCAAATACATATTCA
TCTGACTATTATGTAACACAAATGGAATCAAATTGCAAGGTCATGAAGATTCTATTCT
25 TAGTGTAAAAGTATAGATCAATGATTCTTAATTCTCATCCTCCACGCATAGATCAATGG
TGAATATGGTTTAAATCCTCTAATACTCACTGACTGCCATGGTAGAGTTAAAAAAACAAT
TTTAGAAATATTAGTGGATTAAGGCATTAAGCTGCTCAAGTTGCTTGTATTTC
TTATTAAATTAAAAAAAGTTCAACTATTATTGACTAATAATAACGTGTTAAATGGTTA
TCGGTTAAAATGGCCATAGGCCAGACTGAAGAAAAACTGAAACCCAAAGTTTAT
30 TTTTACTTGTCTTCAGTGAATATCTCCAATCAAGCTTCTCAATTGCTTGCTC
TCTCTCTTACACGGCCAATCGGTGTTCGCAGCTTCAGGTTGTCTCAATCTCAAATTAA
ATCGGAGTCAAGTAATAACAATTGATAACCCTAATTGTTCAATTATATTGAA
ATTTCAGTAGATCCGAATCGTATTCTAGTTCTGGAATCGTTGATCTCGATGGAATT
TTTAAGATTCTTCATACACATTGGTCAAAAGATCACATAATTATTAAATTGATAA
35 GTATGATGATTCTGCTAAGTGGCATTGGATAAAAGTTTCAATTGCAATACGTCTAAACTT
GTCTATGTCTGAATGAACCTCTGAGTTGCTTAAAGTCTTGCTTCTTATTACACA
GGCCTCAATACAAGACATTCTATATAAGCATATTGAGAAGAGGGCGTTCTAATTGTTGC
GGAGTTGAACAATATGACGTAGGGAAATTCTAATTAGGGGAGGCCTCAGAGTTGCACTAA

CTTCATAATCAGCTCTGGACGTTGATTGTATTGAACAAGAACATGTGTAGGCAGAATTGT
CGCGCGAAATCCTCACCGGAGGAAGTGATTCAACTGATGAGAACATCTCTGATATATTGTAA
ACCTGTTGACTATATAACATCTTCACCTTCGCTCTAGGCAACGTATGATTGGCCTTC
CTCTCTCATCATTAGCTTAGTAATCTTCATCTCCTGTGTAGATCACCCACTAATAGTTT
5 GAGTTGCTAAGCTGATTATGGTCTGACTCATGGCAGGTGTGTCTTTGTCTCCTAAT
GTTATTGAACCTGTTGTTGCAGCCATCGTTCTGCCAAGATGCTGAACACTACAA
AATTGGGGCAAAGCGAAAAGAAAGTATGCCTTCTTGAATGTAGTTGCCACAGTGATA
TGTTATTATCTTACTCTAATATGGAAGCTGATGAACATTTATCTTGTGAGTAGATAT
GGACATAATGAATGGTTCTTCTTGTTCATGCTACACTTATATTACAAAATTGTGTT
10 TTGCTTAGGTCAAGACTACTGGATGGTAGTTCAACTATAAGGATTGTAATAATACATT
ACAAAGAACTGAAGGTTAGTCTTTCTGTTCTCGACAAAATTGATGTCAATGTCATGT
TTCTCTAGATGATTGTTACTTTACTATTTCTGTATTGTCACGCAGTTAGGGAGGATTG
TTCTGTCCATTTGCTCTATGCTATGTTGAGCTCAAGGTGGCAACTATTACAACGTGAG
GTTTCTCCGGGGCTTCATATCTAACACTGTGAAATGCTACTGCCGTTAATGCTATATA
15 CTTCACTGTTGGTACATATTTGTGTTGTTGTCTTGTCTCTTAAACTG
CTGAGTGTGTGCTTATCTGAGAAAACATGTTCCAGTTGAGCTTACAATCCATTGTCTTGT
TCTATGCAGGGGCTGCAATTTCATTGAATTCTCATGATTATTTGAATTGAGTTCAA
GGTATGTGGTTTATGGAATTCTTGTGTTGCCTATGCCGTTAGTAATGAGGTTAGTTAA
AAAAGGGTCTTCCTATTGTAGCTTGGAAAGAACACCAGACAGTTAATGTTCTGAAAAC
20 TTAATTCTTCATATTGAGGTCAAGTTACTTAAACTTGGTTAATTGGAAATCCTATAGCT
GGTAAAATTGGTTATATTCCATCCTTATTGTTACTAGGAAGAACAGGAGTGATGATGATA
AATTGAGCCCTCTCTCTGGTAACCTCAGAACCCCTGATTAAATACCTTAATAGCAG
TAACTCCTGCTTTCTGTCACTACTCTCTATAAATCCAACCACAATGTTTGAGCTCG
AAACCTCGTAAGCGTAGACAAAGAGGTGGCAGAAATAACACCAGGAGACTAAAGTATGCTT
25 TTTACCGTTGGATTCAACCAGTTAGCTAATGGCACAGAAAATGGAATTGCCCTGCTGAATG
ATGGTAAAATCACATCTTCTGTGGTATTGTTGTGGCTAGAACCTCATTACAGAACAG
AAGATACAATGTCCTGATTGTTAGTTGTACTCTCCTCGCATTCTTGTGAGGGTA
ATGTTACCAAGAACTGATGTACAAAATTAAATGGCATGCTACAGGAAACCGTGGTTAGGATAT
CCCGAGGCAACAGAGCTGCTGGACAATTGAGATGACTAGCAACATTCCACCAGCCATAGC
30 CCACTCTTCTCTGGACGCTGGCTAAAGTTATATTAAACAACCGAACGCTGTGGTCCCCTGCTA
CTAAGACAAAGAAAGTTATCTGCTGAGCGATCAGAGGCTAGAACGGTTGTTCATCATGACACC
CCGTCATCATAATTACCATACCTGTGTTACAAATGTTCTCCTATTATGGATAAGTGTAA
CTGTACTGCCATATTAAACCGAGAAAATTCTTCAGCCACCTACTCTCAGAAACGCCAAT
TCTATCATTCTCACAGAGTCCAGGTGATCCAAGTTCTCCTCACCTACTTCTTAGGCATTCT
35 TTAAATTGCTCATGATGATATCTTATCAAAGCATACTTGGTTGTTCTCATCTAAATTGTA
TTTGATTCTGTATGTATCAACGCAAAAAATTATGTCATGTTGTCTCCGTTATTGCCA
CTAACCAAAAACAGCATGTTCTGTGACAAGCCAATGGCGCTTGAGCAAGTAATGTCGAT
CGGGATAGCGAGGATGAAGTCGATGACGATGTTGCAGATTGAAAGATGCCAGGTATTCCA

TGATTCTTCTCGTTCATTAAGTAGGCAACAGAAAATGGTACGATGTAACGGCTAAT
GGCTTTGAAACTAAAAAGCTGCAGATGCTGATGACTTGTGGATGTGAATAAGATGA
AAAGCAATTCATGCATCTTGGAACTCGTTGTAAGAAAACAAAGGTAACACTACTCTCTTAC
ACTTGAAACACACACAAAAAGACCTATGCTTACATTCCATACCTGTCTAAATGATTCTGCT
5 TATGGAACCTTGAGCTCAAATTATGATTGATGTTGCAGGGTTAGCAGATGGTCATATCT
CTTGGGCATGTGAAGTATTTCAAGATTTCAGAGAAAGAGTGCAGTGTACTCATCACTC
TTCTGGTAATATAAGTACACCAAACATACAGACACATAACTACACTATCAATTGGTTTC
GTTTTCTGAAAGAAAATAAAAAATTCCAGGTGTTGGAGATTGTTTTGATTAAACTATGG
AACCATGGACTTGTGACTCAGCCACCATAACAACAGTGTGGATCATCCCAGTGACTCAAACACCA
10 TAATACCTCAGTCACTAACAAACAACAAACAGTGTGGATCATCCCAGTGACTCAAACACCA
ACAACAATAACATTGTGGATCATCCGAATGACATAAAAACAAGAACAAATGTTGACAACAAAG
GACAATAACAGCAGAGACAAGTAATTAAATAGGAAACACTCCGGTTAGATGATACCGATCT
ATCGGATTGTAACCTATTCTCTTCTTAAAAAATTGTTAGGAGCAAACAAAGATTTTAT
TTGTTAGTGTATTCAACTGATTACATTAGTTAAAAAATGGATTCTCCTTAATAACTAA
15 AGACTGAAAATAAGATAAGTTCTTAATTTCTTTGACTTGAGAAAAGCTCCTCTA
GACCTCTAGTAAATAGGAGTTATATATTAATCAAGTACATAACATAAAAATATATATTAA
GTGCAAATAGATTGAAAACAAATCAAGAAATTAAAGACACAGTGATTAAGCTTAAACC
CCATTGACTTGTCTTCTCAATGAATCCCTACAAGCAGCAAGCTTGTGATTTGCT
TTGACACCACCAATCAGTGTGTTGAATCTTCAGGTTGTCTGATTTCAAACTAGATCGG
20 AGTCAAGTGATAAAATTGACTAACATAATTATCCGTTCTCGTAAGAGTTGGGATTAG
CAGTAGATCGGAAATCGGAATTACGTTTTGTTAAAAGATTGATGGTTAGGTAAATAAAC
ATAGTTCTGGATTCTTGCTCTAGTTGATTCTCGAATTGTTGATTCGAAATGCACATT
TTGGTTCAAAGGATCACATAATTGCTTAAAATTGACAAAACATACCAAAATTCTCA
TATTCTTCAAGTAACTCGAACCTGTTGGAAATCTATATACTCTGGCTATGTAGATAAGT
25 CTTAACATTGGTCAACATTGTTCTCTAAACTAGTTGGGTTCTGTGTTAAAGTT
TGGTGTCTTCACTATTACACAGGTCTTACAAAGACTACAGTCTCTAGAACATAATCGT
CGACTCTGTTGAGTTCCAACAGTGGTGAAGCTTAAGGAGGTTCTATGTGCCTTTG
AAATC

30

SEQ ID NO: 7

CAAGCTTCTCAATTGCTTGTCTCTCTTACACGGCCAATGGTGTGCGAGCTTT
CAGGCCTCAATACAAGACATTCTATATAAGCATATTGAGAAGAGGGCGGTCTAATTGTTGC
35 ATGGAGTTGAACAATATGACGTAGGGAAATTCTAATTAGGGGAGGCCTCAGAGTTGCAC
AACTTCATAATCAGCTCTGGACGTTGATTGATTGAAACAAGAAATGTGTAGGCAGAATT
GTCGCGCGAAATCCTCACCGGAGGAAGTGATTCAACTGATGAGAATCTCTGATATATTGT
AAACCTGTTGACTATATAACATCTTCACCTCGCTCTAGGCAACCCATGTTCTGCC

AAGATGCTTGAAC TACAAAATTGGGGCAAAGCGAAAAGAAAGTCAGA GATCTACTGGGATGG
TAGTTTCAACTATAAGGATTGTAATAATACATTACAAGAACTGAAGTTAGGGAGGATTGT
TCTTGTCATTTGCTCTATGCTATGTGGTAGCTCAAGGTGGCAACTATTACAAC TGAGG
GGCTGCAATTTCATTTGAATTCATCTCATGATTATTTGAATTGAGTTCAAGCTTTGGAA
5 GAATACCAGACAGTTAATGTTCTGTAAAACCTTAATTCTTCATATTGAGGAAGAAGGAAG
TGATGATGATAAATTGAGCCCTCTCTCTGCTCGAACCTCGTAAGCGTAGACAAAGAG
GTGGCAGAAATAACACCAGGAGACTTAAAGTATGCTTTTACCGTTGGATTCAACCCAGTTA
GCTAATGGCACAGAAAATGGAATTGCCCTGCTGAATGATGGAAACCGTGGTTAGGATATCC
CGAGGCAACAGAGCTTGCTGGACAATTGAGATGACTAGCAACATTCCACCAGCCATAGCCC
10 ACTCTTCTCTGGACGCTGGTGTAAAGTTATTAACAACCGAAGCTGTGGTCCCTGCTACT
AAGACAAGAAAGTTATCTGCTGAGCGATCAGAGGCTAGAACGCCACCTACTTCTCAGAAACG
CCAATTCTATCATTCTCACAGAGTCCAGCCAATGGCGCTTGAGCAAGTAATGTCTGATCGGG
ATAGCGAGGATGAAGTCGATGACGATGTTGCAGATTGAAAGATGCCAGATGCTTGATGAC
TTTGTGGATGTGAATAAAGATGAAAAGCAATTGATCTTGGAACTCGTTGTAAGAAA
15 ACAAAAGGGTTATAGCAGATGGTCATATCTCTTGGCATGTGAAGTATTTCAAGATTTACG
AGAAAGAGTTGCACTGTTACTCATCACTCTTCTGGTGGAGATTGTTGATTAAACTA
TGGAACCATGGACTTGTGACTCAGCCACCATCAACAACTGCAATACCATCCTCGAGAATTG
CCGTAATACCTCAGTCACTAACAAACAACAACAGTGTGGATCATCCCAGTGACTCAAACA
CCAACAACAATAACATTGTGGATCATCGAATGACATAAAAAACAAGAACAAATGTTGACAAC
20 AAGGACAATAACAGCAGAGACAAGTAATTAAATAGGAAACACTCCGGTTAGATGATACCGA
TCTATCGGATTGTAACTTATTCTTCTTCTTAAAAAAATTGTTAGGAGCAAACAAAGATTT
TATTTGTTAGTGTATTCAACTGATTACATTTTAGTTAAAAAAATGGATTCTCCTTAATAAC
T

25

SEQ ID NO: 8

MCRQNCRAKSSPEEVISTDENLLIYCKPVRLYNIFHLRSLGNPSFLPRCLNYKIGAKRKRKS
RSTGMVVFNYKDCNNTLQRTEVREDCSCPFCMSLGSFKVGNYYN

30

SEQ ID NO: 9

ACATTTCGTACCGCTCAAGATTAAGAACCGTAAAAGGGTGGAAATCTCAAGTGATAAAAT
35 TAGGCATGTACATCCACATATTGTGGATTCAAGGATCACCTGAAGATGCCAGGCAGGATCTG
AAGACGATTACGTGCAGAGGGAAAATGGTAGTTCTGTAGCACACGCTCTGTTGATCCTGCT
AATTCAATTACACGGTAGCAATCTTCAGCACCAACAGTGTACAGTTGGGAAGACAAGAAA
GCTGTCTGTTGAACGAGCTGATCCCAGAAATCGGCAGCTCCTACAAAAACGCCAGTTCTTC

ATTCTCACAGGGCTAACCAATGGCATTGGGAGCAGTTCTCAGATCGTATAGTGAAGAT
GAGGTTGATGATGACATTGCTGATTTGAAGATAGACAGATGCTTGATGATTTGTTGATGT
TACCAAAGACGAACCTATTATGCATATGG

5

SEQ ID NO: 10

TFSYRSRFKKRKVEISSDKIRHVPHIVDGSPEDAQAGSEDDYVQRENGSSVAHASVDPA
NSLHGSNLSAPTVLQFGKTRKLSVERADPRNRQLLQKRQFFHSHRAQPMLGAVFSDRDSED
10 EVDDDIADFEDRQMLDDFVDVTKDELIMHM

SEQ ID NO: 11

15 ACATGCATATCCTGATGCTGAATGTGCTCAATTGGTACCTGGAAATAATCTGCACCTCCTG
CCATGCTACAATTCGAAAGACAAGAAAATTATCAATTGAACGGTCTGACATGAGAAACCGT
ACACTCCTTCACAAACGACAATTTTCACTCACATAGAGCTCAGCCAATGGCAGCTGAGCA
AGTTATGTCAGATCGGGATAGTGAGGATGAAGTTGACGATGATGTTGCAGATTTGAAGACC
GAAGGATGCTTGATGATTTGTAGACGTGACTAAAGATGAGAAGCAAATGATGCACTTGTGG
20 AACTCATTGTGAGG

SEQ ID NO: 12

25 HAYPDAECAQLVPGNNLAPPAMLQFAKTRKLSIERSDMRNRTLLHKRQFFHSHRAQPMAAEQ
VMSDRDSEDEVDDDVADEDFEDRRMLDDFVDVTKDEKQMMHLWNSFVR

SEQ ID NO: 13

30 ATGGATCCGATTAAGCTGACAACAGAACAGCTAAGGTCCCTGCTAACGATCAAAGGCTACAAG
CCACTACTTGCCTCTTCATAAACGCCAGTTCTATCATTCCGAACCGGTCAGCCATTGTCAC
TTGAGCAAGTTATGTCGACCGAGATAGCGAAAATGACGTCGACAAAAATGATGATGCTGCA
CATCTCGAAGAAAGCCAGATGCTTAATGGTCCATGGATGAGAATGAAATCGTAGCAGAGAG
ATTCAATAAAACTTGGAACTCCTTGTTAACAGCAAAGGATTGTTGCAGATGCTCATATTG
35 CTTGGGCATGTGAAGCATTCTCAAGATTACACCTGCAAGAGCTGCGCAGTAACCTTACACTC
GACTTGTGCTGGAGACAATTATGATCAAACAAATGGGATTATGGACTTCTGACAGAGTCAC
CATGAACAAATGCAATACCATCATCTACCAATAATCTCAACTACCAACGATGACATAAACAA
ATAACAAACACAAGGACGACTGATAATATGGATGTTGTCGACGATGACATAAACAGAGACAAG

SEQ ID NO: 14

MDPIKLTTEAKVPAKRSKATSHYLPLHKRQFYHSRTGQPLSLEQVMSDRDSENDVDKNDDAA
5 HLEESQMLNGSMDENEIVAERFIKLWNSFVKQQRIVADAHIPWACEAFSRLHLQELRSNLSL
DLCWRQFMIKQWDYGLLDRVTMNKNTI IYHNISTTNDDINNNNTRTTDNMDVVDDDINRDK

SEQ ID NO: 15

10 CTCTGAGGAGACACTTTTTCCCTCCCTCCTCCCTCTCCCTCCCTCCCTCCCTCCCTCC
CTCTCCTCCCCTCTCCTCCTCCCCCTCGTCCGCCGGAGCCTGCTGGGGCGAGCGGTT
GGTATTGCAGGCGCTTGCTCTCGGGGCCCGCCGGTAGCTGGCGGGGGAGGAGGCAG
GAACCGCGATGGCGCCTCAGAACGACGGCGGTGGGGAGGGGGCGGCTCGGGGCCAGCGC
15 GGGTCGGGGAGGCGGCTCGGGGTTCGGCGGGTGGCGGGCGACGGCTCGGGCGG
CAAATCCGGCGCGGGAGCTGTGGAGGGGGTGGCAGTTACTCGGCCTCCTCCTCCTCCG
CGGCAGCAGCGGGCTCGGTGTTACCGGTGAAGAACGAAAATGGAGCACGTCCAG
GCTGACCACGAGCTTCCTCCAGGCCTTGAGAACGAAACACAGATCTATAGATTCTCG
AACTCGGAATCTCATAGCACCAATATTTTGACAGAACTCTTACTTACATGTCTATCGA
20 ACTCCAGAACAAACATCAAAAGAAAACATTAAAGTTGATGATATGTTATCAAAAGTAGAG
AAAATGAAAGGAGAGCAAGAACATCTCATAGCTTGTCACTTCAGCTTACGTTACTGG
TTTCTCCACAAAAATGATAAGCCATCACAAACTCAGAAAATGAACAAAATTCTGTTACCC
TGGAAAGTCCTGTTGTGAAAGTTGCCACAAAAAGAAAGGATGTAAGTTGTCCAATAAGG
CAAGTCCCACAGGTAAAAGCAGGTGCCTTGATTCTGACCTCAATCAAACAAAACCCGG
25 AAATTCCCGTCCCTGCAGTTCCAGTAATGAATTGAAACCTAGTAACAGCCATATGGTGA
AGTCTTACTCGTTGCTATTAGAGTGACTCGTCCAGGAAGAACAGAGTTAATGGAATGATT
AATGGAGAAACCAATGAAAATATTGATGTCAATGAAGAGCTTCCAGCCAGAACAGAAA
TCGTGAGGATGGGGAAAAGACATTGTTGCACAAATGACAGTATTGATAAAAACAGCGCT
TACAGCTTTAGATGGGAATATGAAGTAGCCATGCAGGAAATGGAAGAACAGTCCAATAAGC
30 AAGAAAAGAGCAACATGGGAGACTATTCTGATGGGAAGAGGCTGCCTCCATTGAAACATT
TTCTCAGGGACCTACGTTGCAGTTCACTCTCGTTGGACAGGAGAGACATGATAATCTA
CGGCTCCTATTGCCAAACCTCTGCCACTAGAAATTAGAGAGCTCCATCAGGAAAACAAG
CCTGGTTCACTTAAACACTCAAATGAAACCGACAAAATTAAGAACATTTTATCAGT
AACAAAGAAAAGAAAAGGATACTCCAAATGAAACCGACAAAATTAAGAACATTTTATCAGT
35 TTCTCTATAACAACAATACAAGGCAACAAACTGAAGCAAGAGATGACCTGCATTGCCCTGG
TGTACTCTGAACTGCCGAAACATTATAGTTACTCAAGCATCTTAAACTCTGCCATAGCAG
ATTTATCTCAACTATGTTATCATCCAAAAGGTGCTAGGATAGATGTTCTATCAATGAGT
GTTATGATGGCTCCTATGCAGGAAATCCTCAGGATATTCATGCCAACCTGGATTGCTTT

AGTCGCAACGGACCAGTTAAGAGAACACCTATCACACATATTCTTGTGTGCAGGCCAAACG
AACAAAAGCAAGCATGTCTGAATTCTTGAATCTGAAGATGGGAAGTAGAACAGCAAAGAA
CATATAGTAGTGGCCACAATCGTCTGTATTCCATAGTGATACTGCTTACCTCTCCGTCCA
CAAGAAATGGAAGTAGATAGTGAAGATGAAAGGATCCTGAATGGCTAAGAGAAAAACCAT
5 TACACAAATTGAAGAGTTTCTGATGTTAATGAAGGAGAGAAAGAAGTGTGAAACTCTGGA
ATCTCCATGTCTGAAGCATGGTTATTGCTGACAATGAATCATGCCTGTATGCTG
TTTGTAGAAAATTATGGACAGAAAATAATTAGAAGAATTATGTCGAAACTTCATGCTTCA
TCTAGTCAGCATGCATGACTTAATCTTATTAGCATAATGTCATAAGATAAAGCTGTTACCA
AGCTCCGTGAAATGCAGCAAAATTAGAAAAGGGGAATCTGCTCCCTGCAAACGAAGAA
10 ATAACGTGAAAGAACAAATGGACAGCAATGGATTAGTGAAATTAACTCAAAAGAGAAAGC
TTTGGAAACAGATAGTGTCTCAGGGTTCAAAACAGAGCAAAAACAAAAACTCTGAAAAG
CTCTAACCCCATGTTATGGACAAACACTGAAATTACATTAGGAAATTCACTCTAAGAA
TTATGTTTTGTTTAATCATATGTTCAAACAGGCAGTGTAGATGAAAGTAAATGATTTC
AACAGGATATTGTATCAGGTTCTACTTCATTATGCAGCATTACATGTATATCAC
15 TTTTATTGATGTCATTAAACATTCTGTACTTTAAGCATGAAAAGCAATATTCAAAGTATT
TTTAAACTCAACAAATGTCATCAAATATGTTGAATTGATCTAGAAATTATTCATATATAAA
TCAGAATTTTTGCATTTGACGGCTGTTCTACTTGTAAATTGTGAGACATTTC
TGGGGAGGGAAAATTGGAATGGTCCCTTTAGAAATTGAAAGTGGTCTTCATATGTCAAC
TACAGAAAAGAAAAAATAGAAATTGAGGATTTTATGAAATTATATTGCATTACTATT
20 GCAGTCAAACTTGATCCTGTTTGAAATCATTGTCATTGCAATTGGAATGAAAATTATAAT
GTAATTTCACATTACATAAGTCCCTTACAATTAAAAAATAGCACTTCTTCATCTTATGCC
TGTTGAGAAGATATTAAATTTCACATTGTTGACAGTGAAATGCTATGTTGGTTATAAGA
TTACAGACCATTGTTTCATGTGGATAATTTCAGTGCTCACCCTGTT
TTTTTAACTTGAACATTGCTGTTGTTCTTTAATTAGATAATCACACGGAAA
25 ATTAAGCTGTTCATATCTTAAATTAGGATTGCAAACCAAGGAAAGAACGCAATTGAGATT
TAAGATGTCACTTATAAGGGAGAAGTGTCTTAAAAGTCACCAAGGAAACTGTTATGCCT
TTTATTGTTGCAAGGATGTCTTGTAAATGTGTTCATGAATAGAATATCCAATAGAGATA
AGCTGACTTGAATCATTGAGCAATTGCCCCGTGTTATATGTGTTCACGCACATATT
GCAGTTGGATTTCCTCAAACAGAAAGTGGATTCACTACTGGCACATTAACAAGCACCAATAG
30 GTTTTATTCCAACCTCGAGCACTGTGGTTGAGTAACATCACCTCAATTTCATTATCCTT
AAAGATATTGCAATTGTTCATATTCTTATTATAAAGGATCAATGCTGCTGAAATACAGGTA
TTTTTAATTAAAATTCCATTCCACCACCATCAGATGCAGTCCCTATTGTTAATGAA
GGGATATATAAGCTTCAATGGTGTCTCAGAAATTATAAAATGAAATTACTGATTGAC
TGGCTTTAAGATGTGTTAAGTGAGGCTATTAAACGAATAGTGTGGATGTGATTGTC
35 TCCAGTATTAAAGTTCTTAGTCATTGATTGTTGTTAAAAAAAAAATAGGAAAGAGGGAAAC
TGCAGCTTCATTACAGATTGCTTGATTGTAAGCTCTCCAAATGATGAGTCTAGTAAACT
CTGATTTCGCTCTGGATAGTAGATCTCGAGCGTTATCTGGGCTTTAATTGCTAAAGC
TGTGCACATATGAAAAAAAAAAGATTATTTAGGGAGATGTAGGTGAGAATT

ATTGCTTATGTCATTCTTAAGCAGTTATGCTCTTAATGCTAAAAGAAGGCTAGCATTGTT
TGCACAAAAAGTTGGTGATTCCCACCCAAATAGTAATAAAATTACTCTGTTGAGTAAACT
TTTTATGTCATCGTAAAAGCTGGAAAATCCCTTGTTCATTATAAAAAAGTGTCTTT
CTATATGTACCCTTGATAACAGATTTGAAGAAATCCTGTAAGATGATAAAGCATTGAATG
5 GTACAGTAGATGTAaaaaaaaATTCAAGTTAAAAGAACATTGTTTACATTAATGTTAT
TTGAAATCAAATGATTTGTACATAAAGTTCAATAATAT

SEQ ID NO: 16

10

LRRHFFPPSFPLLLPSLPLSSPLSSFPPRSAGACWGERLVLQALALRGRPAGSWRGEEAG
TAMAPQKHGGGGGGSGPSAGSGGGFGGSAAVAAATASGGKSGGGSCGGGSYSASSSSSA
AAAAGAAVLPVKPKMEHVQADHEFLQAFEKPTQIYRFLRTRNLIAPIFLHRTLTYMSHRN
SRTNIKRKTFKVDDMLSKEVMKGEQESHLSAHLQLTFTGFFHKNDKPSPNSENEQNSVTL
15 EVLLVKVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTKPGNFPSSLAVSSNEFEPNSHMK
SYSLLFRVTRPGRREFNGMINGETNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRL
QLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPPFETFSQGPTLQFTLRWTGETNDKST
APIAKPLATRNSESILHQENKPGSVKPTQTIAVKESLTDLQTRKEKDTPNENRQKLRIFYQF
LYNNNTRQQTEARDDLHCPWCTLNCRKLYSLLKHLKLCNSRFIFNYVYHPKGARIDVSINEC
20 YDGSYAGNPQDIHRQPGFAFSRNGPVKRTPIHILVCRPKRTKASMSEFLESEDGEVEQQRT
YSSGHNRLYFHSDTCLPLRPQEMEVDSDEDEKDPEWLREKTITQIEEFSDVNEGEKEVMKLWN
LHVMKHGFIAIDNQMNHACMLFVENYGQKIIKKNLCRNFLHLVSMHDFNLISIMSIDKAVTK
LREMQQKLEKGESASPAANEEITEEQNGTANGFSEINSKEALETDSVSGVSKQSKKQL

CLAIMS:

1. An isolated nucleic acid obtainable from the VRN2 locus of a plant, which nucleic acid encodes a polypeptide which is capable of affecting one or more physical characteristics of a plant into which the nucleic acid is introduced, the physical characteristics being selected from vernalization response, flowering time, leaf size and/or shape or shade avoidance response.
- 10 2. A nucleic acid according to claim 1 which is capable of reducing the vernalization requirement of the plant for flowering.
- 15 3. A nucleic acid according to claim 1 or claim 2 which comprises a VRN2 nucleotide sequence which encodes a polypeptide of SEQ ID No. 2.
- 20 4. A nucleic acid according to claim 3 wherein the VRN2 nucleotide sequence consists of the sequence of SEQ ID No. 1 or is degeneratively equivalent thereto.
- 25 5. A nucleic acid according to claim 1 or claim 2 which comprises a VRN2 nucleotide sequence which encodes a polypeptide of SEQ ID No. 5.
- 30 6. A nucleic acid according to claim 5 wherein the VRN2 nucleotide sequence consists of the sequence of SEQ ID No. 4 or is degeneratively equivalent thereto.
7. A nucleic acid according to claim 1 or claim 2 which comprises a VRN2 nucleotide sequence which encodes a polypeptide of SEQ ID No. 8.
- 35 8. A nucleic acid according to claim 7 wherein the VRN2 nucleotide sequence consists of the sequence of SEQ ID No. 7 or is degeneratively equivalent thereto.

9. An isolated nucleic acid obtainable from the VRN2 locus of a plant which nucleic acid comprises a nucleotide sequence of SEQ ID No. 3 or SEQ ID No. 6, or a nucleic acid degenerative equivalent thereof.

5

10. A nucleic acid according to claim 9 which comprises a sequence having promoter and/or regulatory function.

11. An isolated nucleic acid which comprises a variant
10 sequence which is a homologous variant of a nucleotide sequence of any one of claims 4, 6 , 8, 9 or 10 and which shares at least about 50% identity therewith.

12. A nucleic acid according to claim 11 wherein the variant
15 sequence is capable of affecting one or more physical characteristics of a plant into which the nucleic acid is introduced, the physical characteristics being selected from vernalization response, flowering time, leaf size and/or shape or shade avoidance response

20

13. A nucleic acid according to claim 11 or claim 12 wherein the variant is an allelic variant of a nucleotide sequence of any one of claims 4, 6, 8, 9 or 10, or is a nucleic acid degenerative equivalent thereof.

25

14. A nucleic acid according to claim 11 or claim 12 wherein the variant sequence is a VRN2 sequence obtainable from a plant species other than *Arabidopsis thaliana*.

30 15. An isolated nucleic acid sequence which comprises a derivative of a nucleotide sequence of any one of claims 4, 6, 8, 9 or 10, which is a derivative by way of one or more of addition, insertion, deletion or substitution of a said nucleotide sequence and wherein the derivative sequence is
35 either capable of affecting one or more physical characteristics of a plant into which the nucleic acid is introduced, the physical characteristics being selected from vernalization response, flowering time, leaf size and/or shape

or shade avoidance response or has promoter and/or regulatory function.

16. An isolated nucleic acid which comprises a sequence which
5 is a fragment of a sequence of any one of claims 1 to 15 and
either is capable of affecting one or more physical
characteristics of a plant into which the nucleic acid is
introduced, the physical characteristics being selected from
vernalization response, flowering time, leaf size and/or shape
10 or shade avoidance response or has promoter and/or regulatory
function.

17. An isolated nucleic acid which comprises a sequence which
is the complement of a sequence of any one of claims 1 to 16.
15

18. An isolated nucleic acid for use as a probe or primer
which comprises a sequence that encodes an amino acid sequence
that is partly, substantially or completely conserved between
a VRN2 sequence of SEQ ID Nos. 2, 5 or 8 and at least one of
20 the other sequences shown in Figure 8a or 8b.

19. An isolated nucleic acid for use as a probe or primer
which comprises a sequence that is partly, substantially or
completely conserved between two or more of the VRN2
25 nucleotide sequences of claim 4, 6, 8, 9 or 10 or the
complements thereof.

20. A nucleic acid according to claim 18 or claim 19 which is
15 to 40 nucleotides in length.
30

21. A nucleic acid according to claim 20 which is 19 to 28
nucleotides in length.

22. A pair of primers each comprising a nucleic acid
35 according to any one of claims 18 to 21.

23. A pair of primers according to claim 22 which is selected from:

VRN2-AP and VRN2-AJ;
VRN2-AO and VRN2-AS; and
5 VRN2-AI and VRN2-AJ.

24. A process for producing a nucleic acid which is a derivative according to claim 15 which process comprises the step of modifying

10

25. A method for identifying or cloning a nucleic acid according to any one of claims 1 to 17, which method comprises using a nucleotide sequence of a probe or primer of claim 18 15 or 19 in a database search.

20

26. A method for identifying or cloning a nucleic acid according to any one of claims 1 to 17, which method employs a probe or primer according to any one of claims 18 to 21 or a pair of primers according to claim 22 or 23.

25

27. A method for determining the presence of a nucleic acid according to any one of claims 1 to 17 within the genetic make-up of a plant, which method employs a probe or primer according to any one of claims 18 to 21 or a pair of primers 30 according to claim 22 or 23.

28. A method according to claim 26 or claim 27, which comprises the steps of:

35

- (a) providing a preparation of nucleic acid from a plant cell;
- (b) providing a nucleic acid molecule which is a probe according to any one of claims 19 to 21;
- (c) contacting nucleic acid in said preparation with said probe under conditions for hybridisation; and
- (d) identifying a nucleic acid according to any one of claims 1 to 17 if present by its hybridisation with said nucleic acid probe.

29. A method according to claim 26 or claim 27, which method comprises the steps of:

(a) providing a preparation of nucleic acid from a plant cell;

5 (b) providing a pair of nucleic acid molecule primers suitable for PCR, at least one of said primers being a primer of any one of claims 18 to 21;

(c) contacting nucleic acid in said preparation with said primers under conditions for performance of PCR;

10 (d) performing PCR and determining the presence or absence of an amplified PCR product.

30. A method according to claim 29 wherein the pair of nucleic acid molecule primers are according to claim 22 or
15 claim 23.

31. A method of selecting a plant having a desired allele of the VRN2 gene, which method employs a probe or primer according to any one of claims 18 to 21 or a pair of primers
20 according to claim 22 or 23.

32. A recombinant vector which comprises the nucleic acid of any one of claims 1 to 17.

25 33. A vector according to claim 32 wherein the nucleic acid comprised in the vector is capable of regulating one or more genes involved in the transition from vegetative to reproductive growth and/or capable of regulating one or more genes involved in the determination of leaf size and/or shape.
30

34. A vector according to claim 32 or claim 33 wherein the nucleic acid is operably linked to a promoter for transcription in a host cell, wherein the promoter is optionally an inducible promoter.

35 35. A vector according to claim 34 wherein the promoter is a constitutive promoter.

36. A vector according to any one of claims 32 to 35 which is a plant vector.

37. A method which comprises the step of introducing a vector
5 according to any one of claims 32 to 36 into a host cell such as to transform the host cell.

38. A host cell which is transformed with a heterologous nucleic acid of any one of claims 1 to 17.

10

39. A host cell according to claim 38 which is a plant cell, optionally present in a plant.

40. A method for producing a transgenic plant, which method
15 comprises the steps of:

- (a) performing a method according to claim 37
- (b) regenerating a plant from the transformed host cell.

41. A transgenic plant which is obtainable by the method of
20 claim 40, or which is a clone, or selfed or hybrid progeny or other descendant of said transgenic plant, which in each case includes the plant cell of claim 39.

42. A plant according to claim 41 which is an agricultural or
25 horticultural plant.

43. A plant according to claim 41 or claim 42 selected from the list consisting of: rice, maize, wheat, barley, oats, rye, oil seed rape, sugar beet, sunflower, soybean, sorghum,
30 lettuce, endive, cabbage, broccoli, cauliflower, carnation, geranium, tobacco, cotton, canola, tomato, mango, peach, apple, pear, strawberry, banana, melon, carrot, onion, pea, celery.

35 44. A plant according to claim 43 which is selected from tobacco, oil seed rape, rice and wheat.

45. A part or propagule from a plant according to any one of claims 41 to 44, which includes the plant cell of claim 39.

46. An isolated polypeptide which is encoded by a nucleotide sequence according to any one of claims 1 to 17.

47. A polypeptide according to claim 46 which comprises an amino acid sequence which consists of the sequence of SEQ ID Nos. 2, 5 or 8.

10

48. A polypeptide according to claim 46 which is a fragment of a polypeptide of claim 47.

49. An isolated polypeptide which consists of the sequence of SEQ ID Nos. 2, 5 or 8.

50. An isolated polypeptide which is a fragment of a polypeptide according to claim 47 and which comprises at least 5 contiguous amino acids of that polypeptide.

20

51. A fragment according to claim 50 which comprises amino acids 90 to 111 (zinc finger motif), amino acids 63 to 132 (amino terminal region), amino acids 263 to 366 (carboxy terminus) or amino acids 263 to 328 (activation domain).

25

52. A fragment according to any one of claims 48 to 51 which is capable of affecting one or more physical characteristics of a plant expressing said fragment, the physical characteristics being selected from vernalization response, flowering time, leaf size and/or shape or shade avoidance response.

53. An isolated nucleic acid which encodes for a fragment according to claim 50 or claim 51, or claim 52 as dependent therefrom.

54. An isolated nucleic acid which is the complement of the nucleic acid of claim 53.

55. A method of making the polypeptide of any one of claims 46 to 52, which method comprises the step of causing or allowing expression from a nucleic acid of any one of claims 1 to 17 or 53 in a suitable host cell.

5

56. An antibody which has specific binding affinity for a polypeptide according to any one of claims 46 to 52.

57. A polypeptide which comprises the antigen-binding site of 10 the antibody of claim 56.

58. A method for affecting a physical characteristic of a plant selected from vernalization response, flowering time, leaf size and/or shape or shade avoidance response, which 15 method comprises either the step of:

- (i) causing or allowing transcription from a nucleic acid according to claim 17 or claim 54 in the plant; or
- (ii) causing or allowing transcription from a nucleic acid according to any one of claims 1 to 16 or claim 53.

20

59. A method according to claim 58 which method comprises causing or allowing transcription from a nucleic acid according to any one of claims 1 to 16 or claim 53 thereby to reduce VRN2 expression by co-suppression.

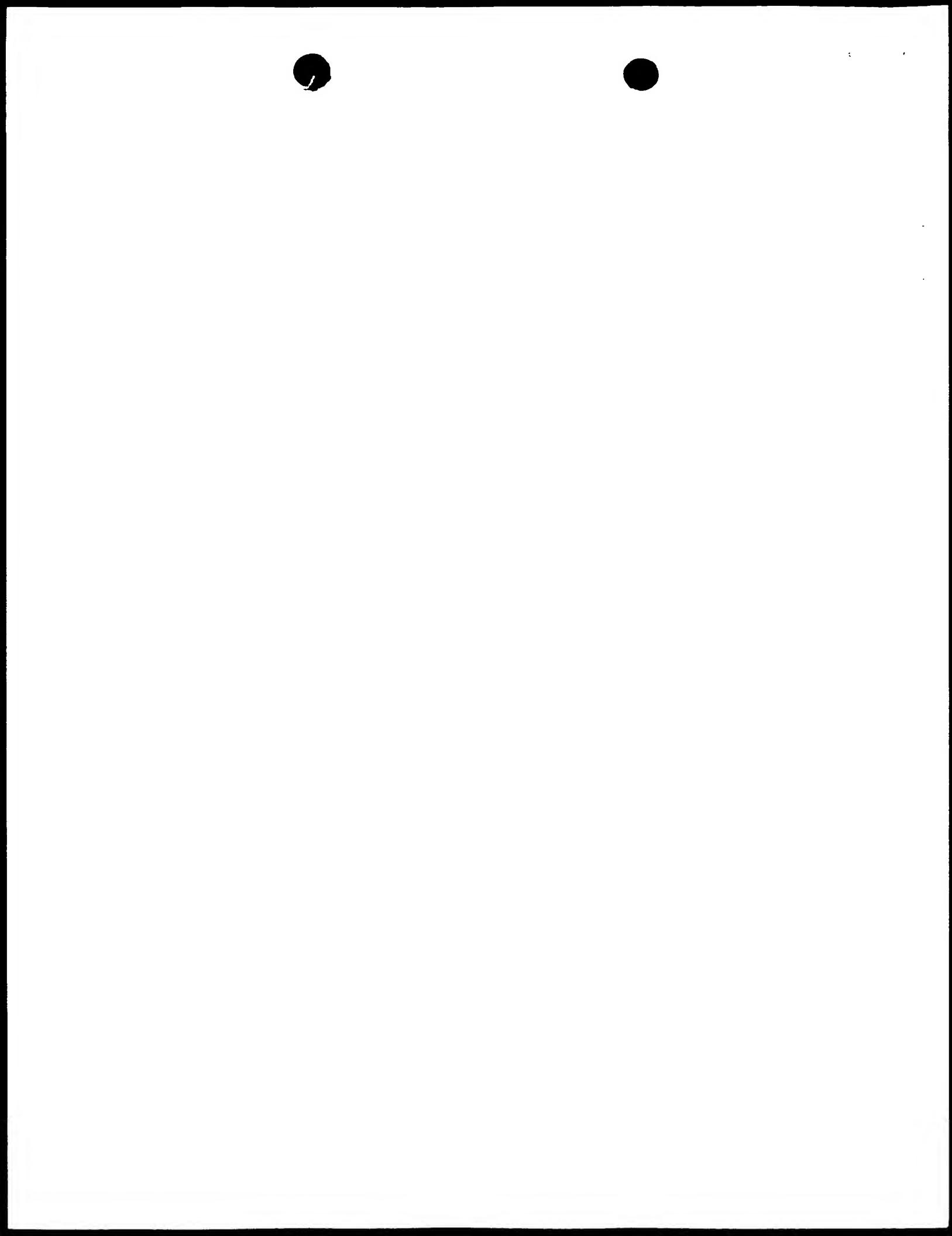
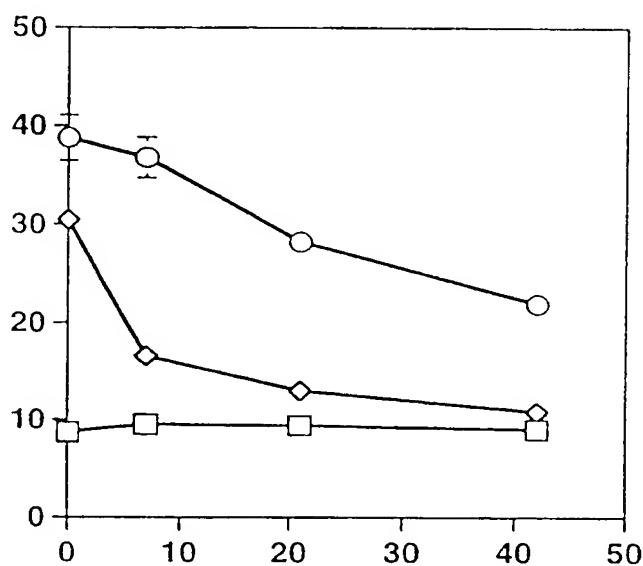
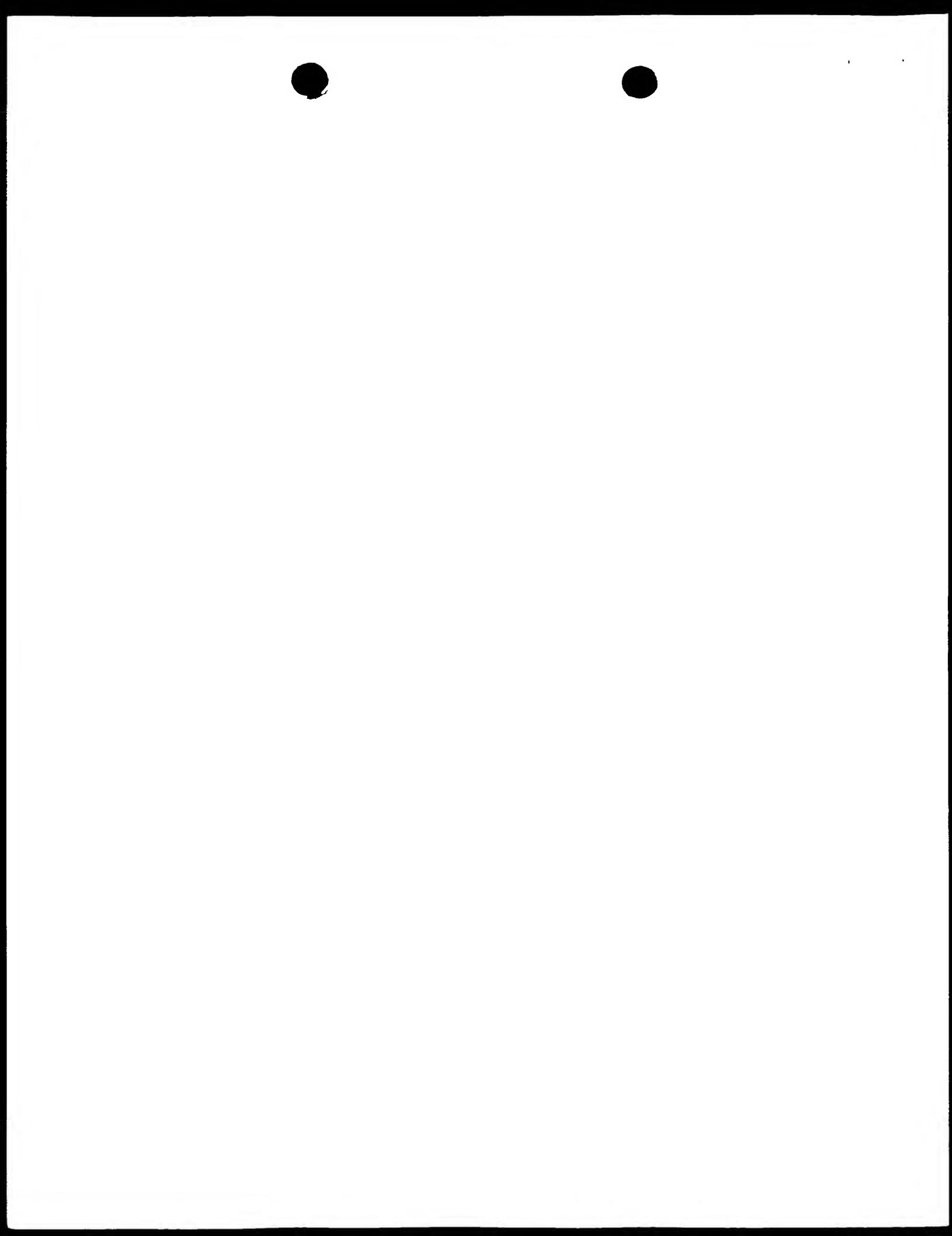
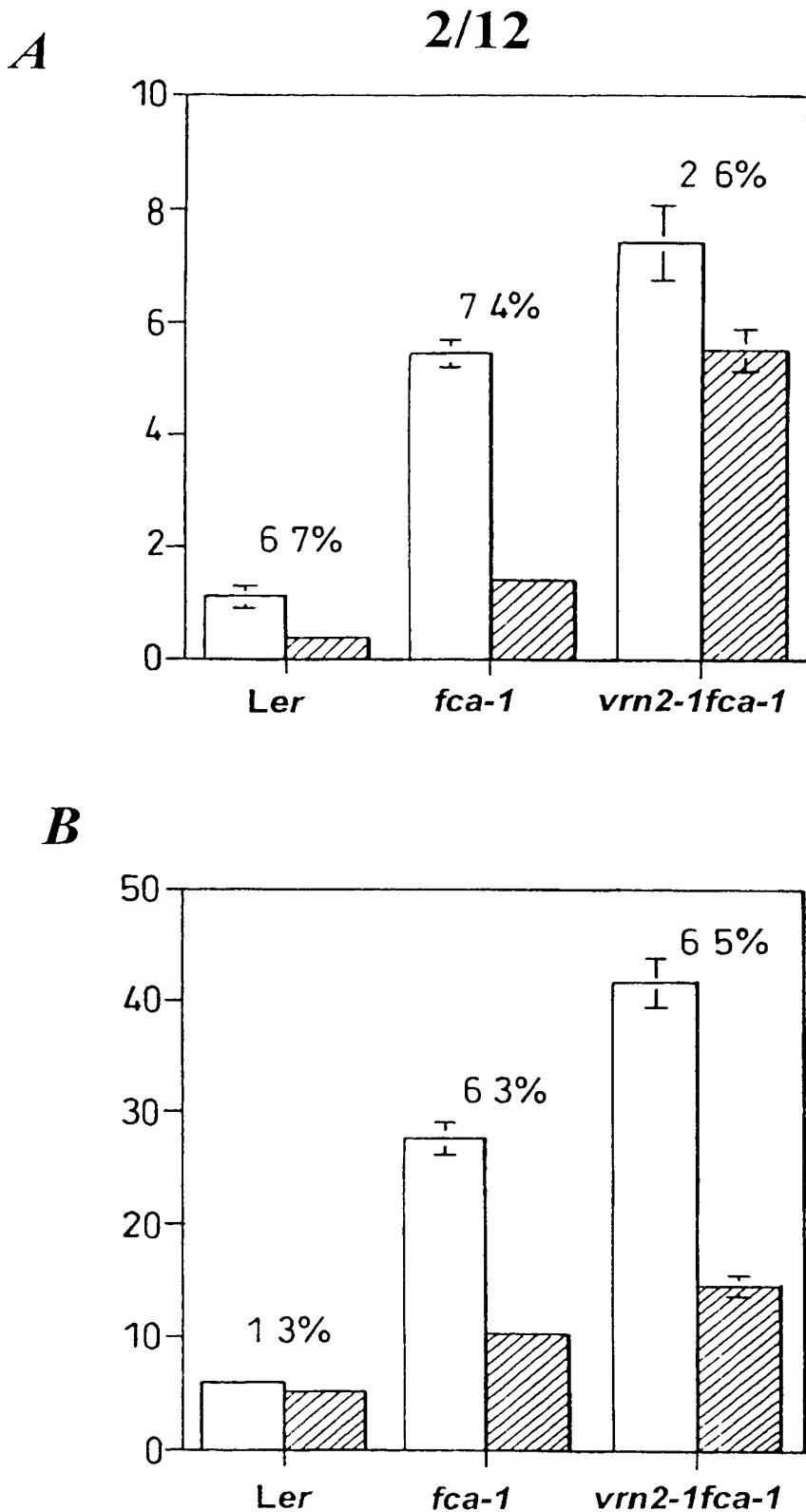


Figure 1



**Fig. 2**

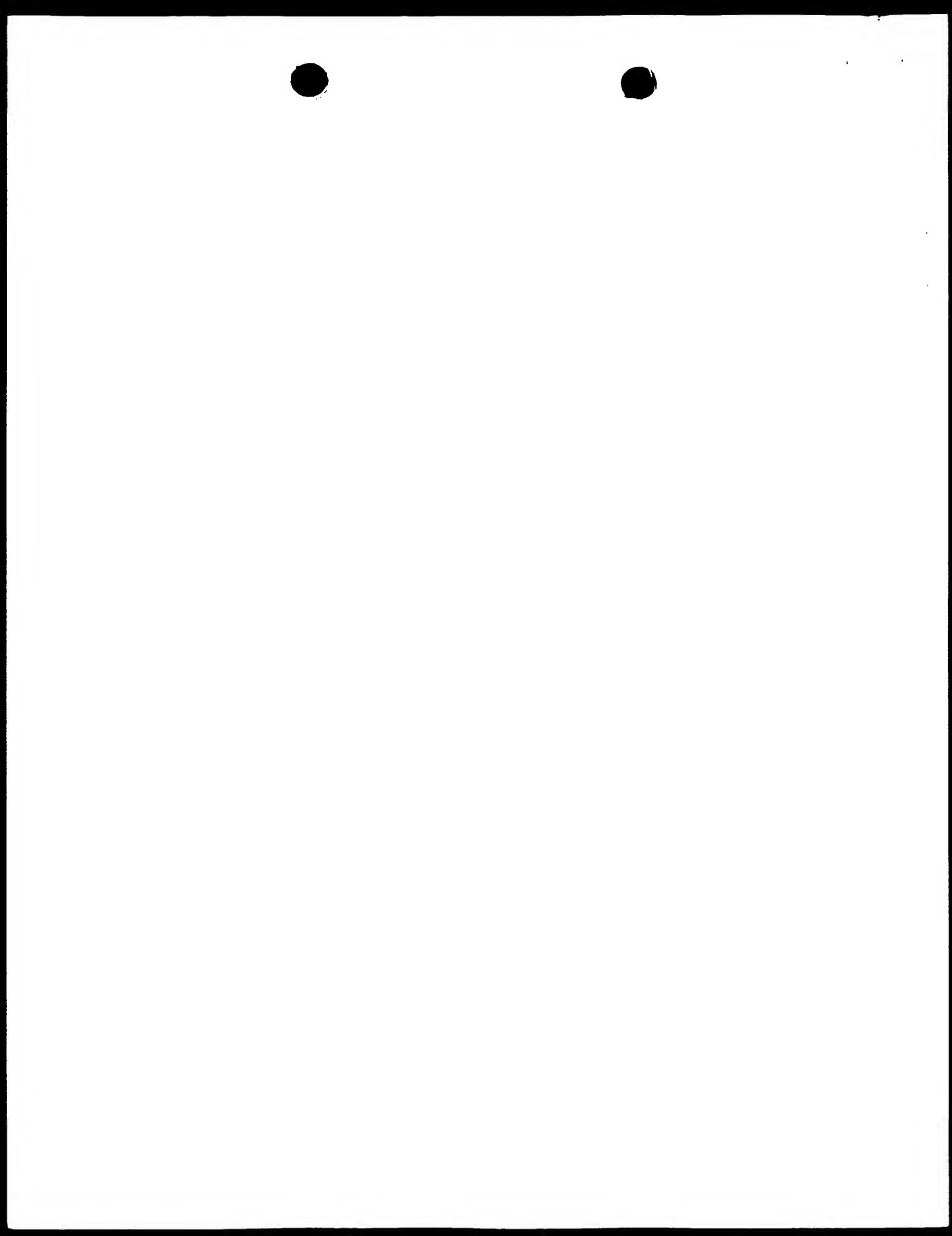
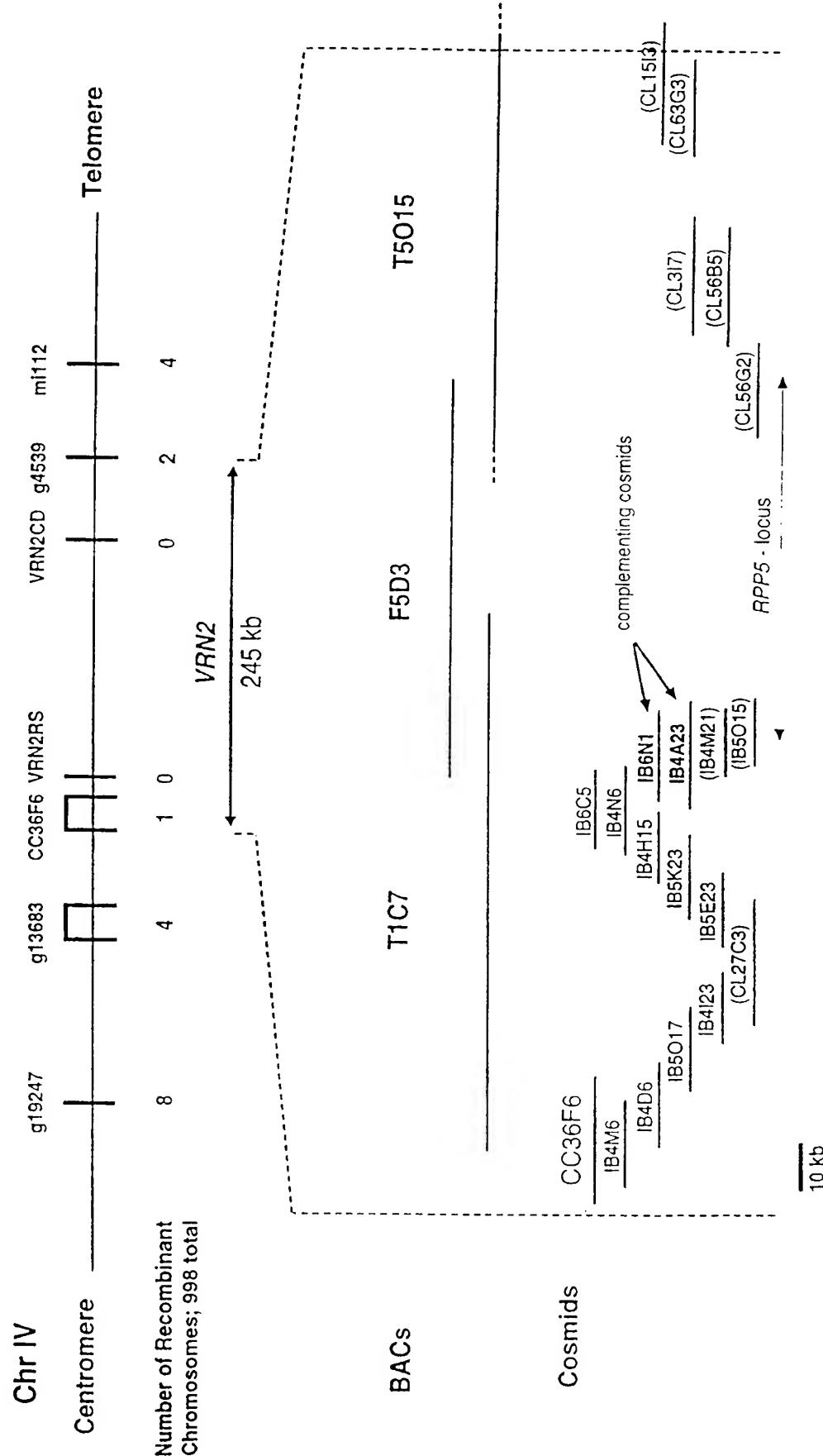
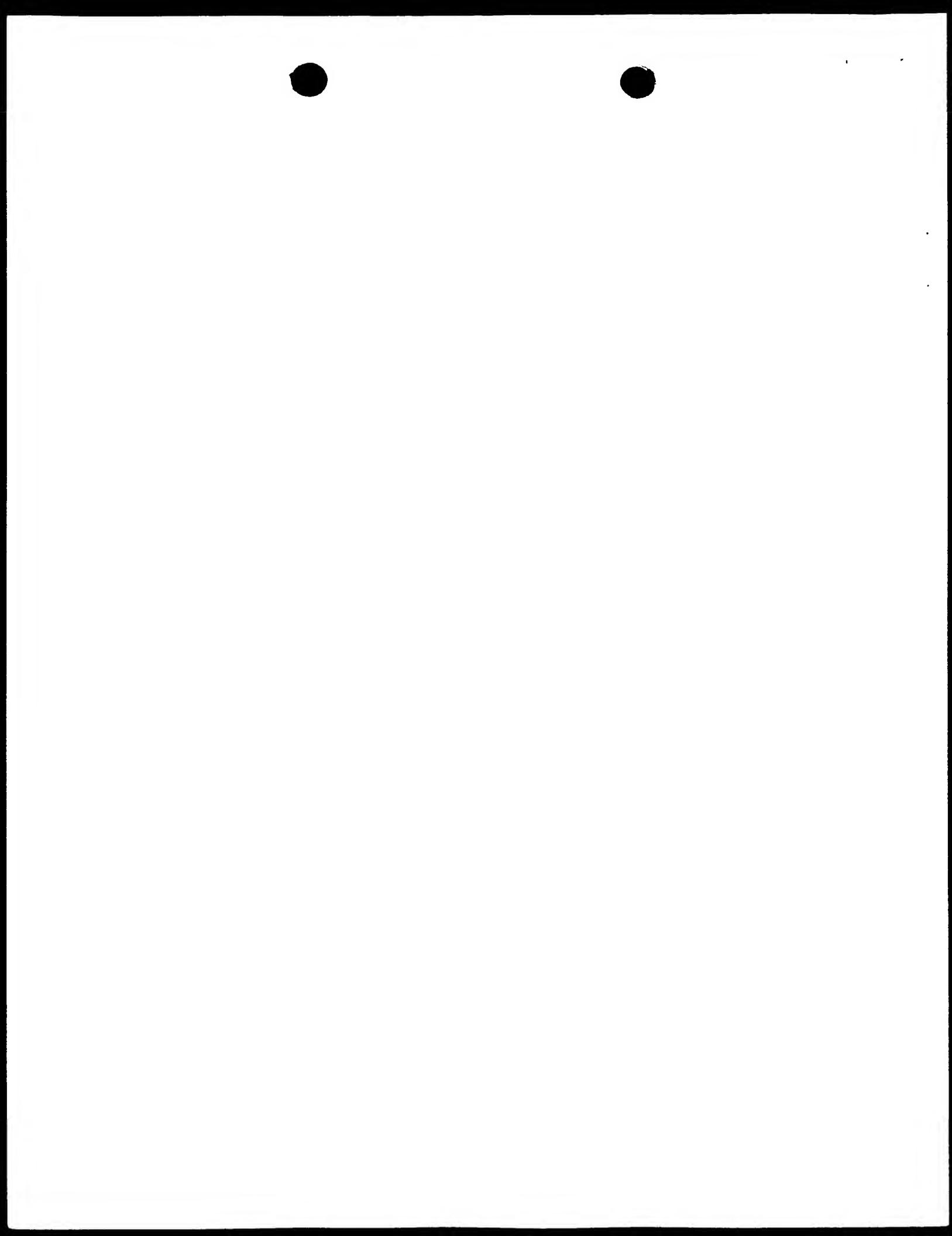


Figure 3





4/12

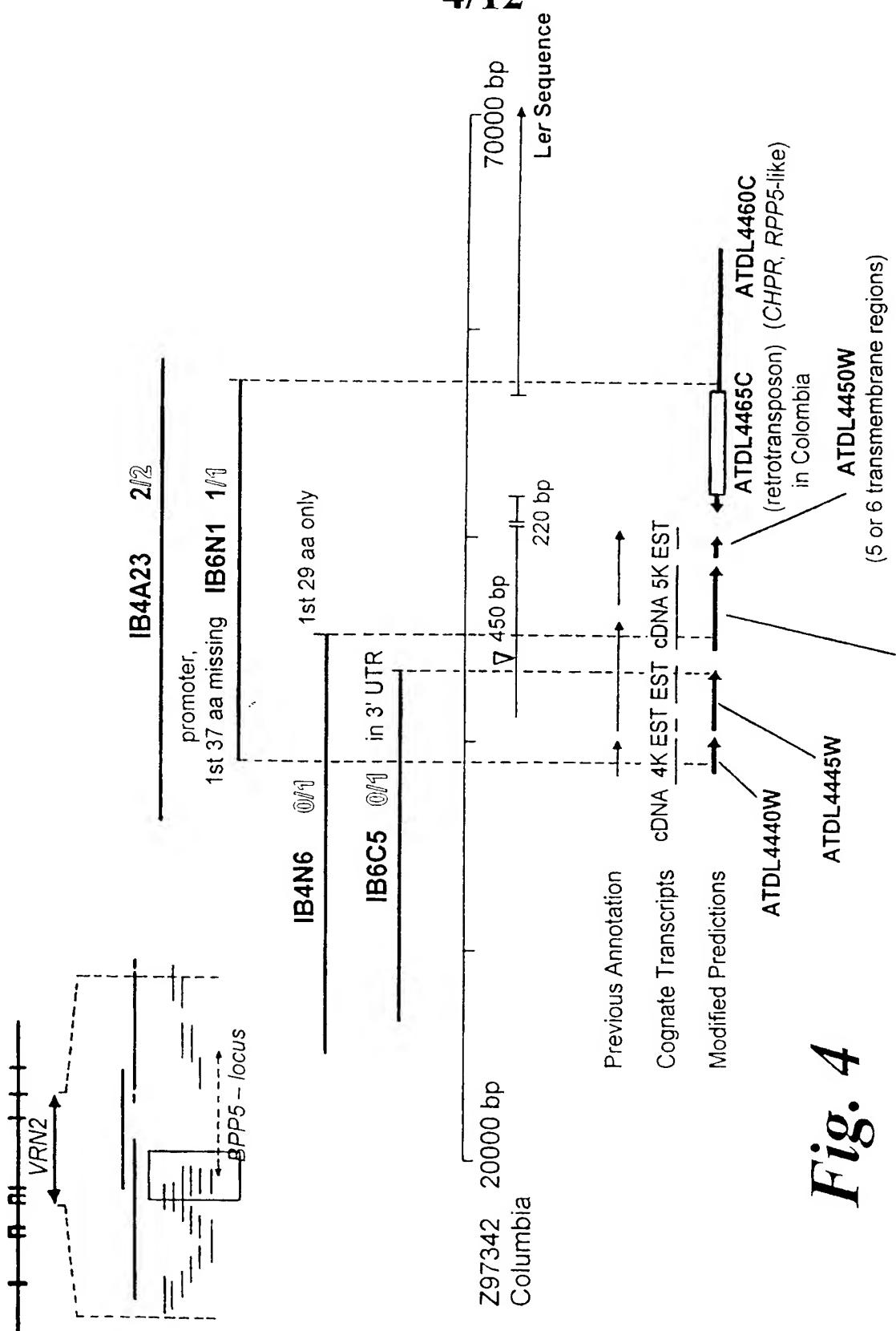


Fig. 4

5K unknown, 15 exons
 (predicted Arabidopsis protein,
 Rice EST, human and mouse ES)

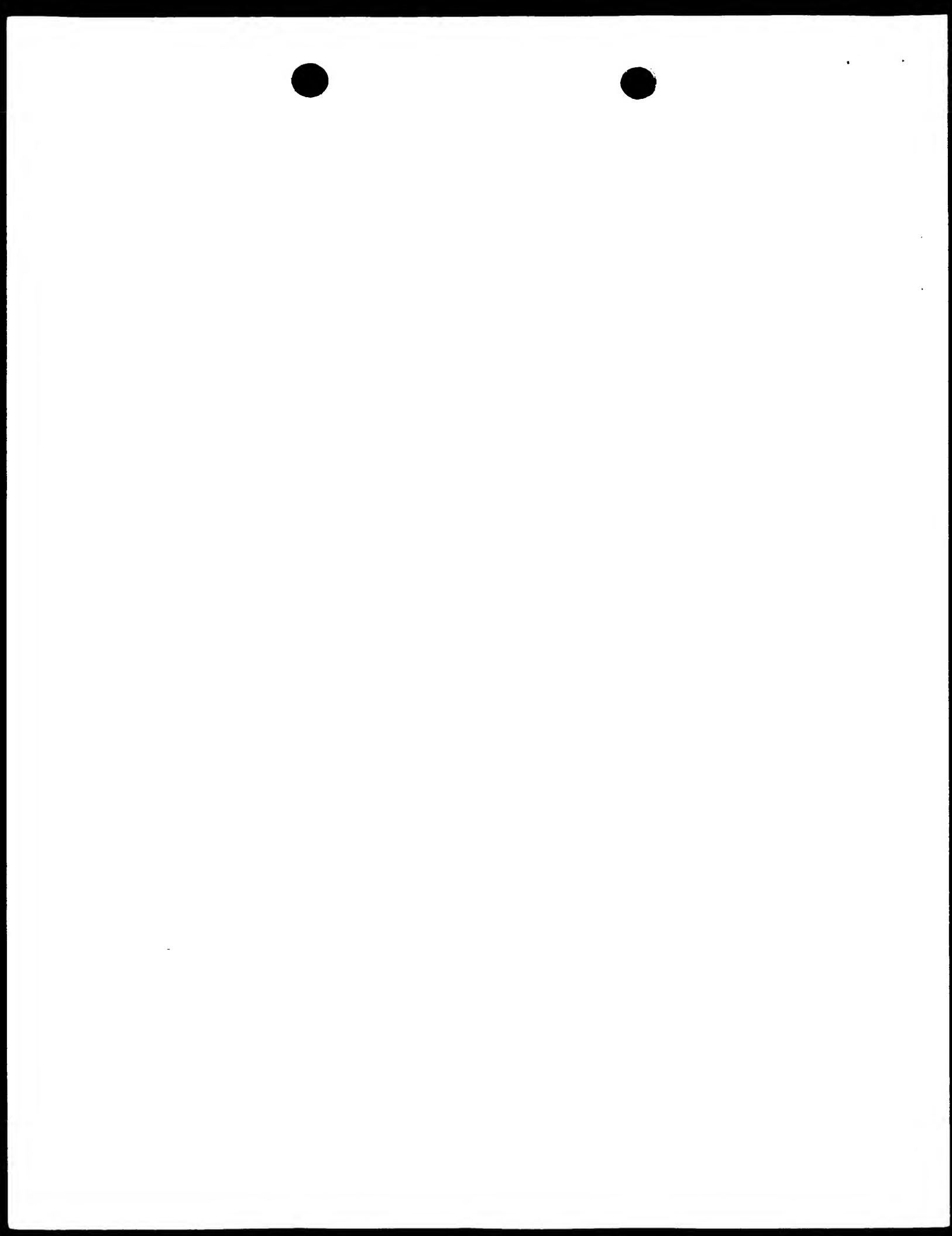
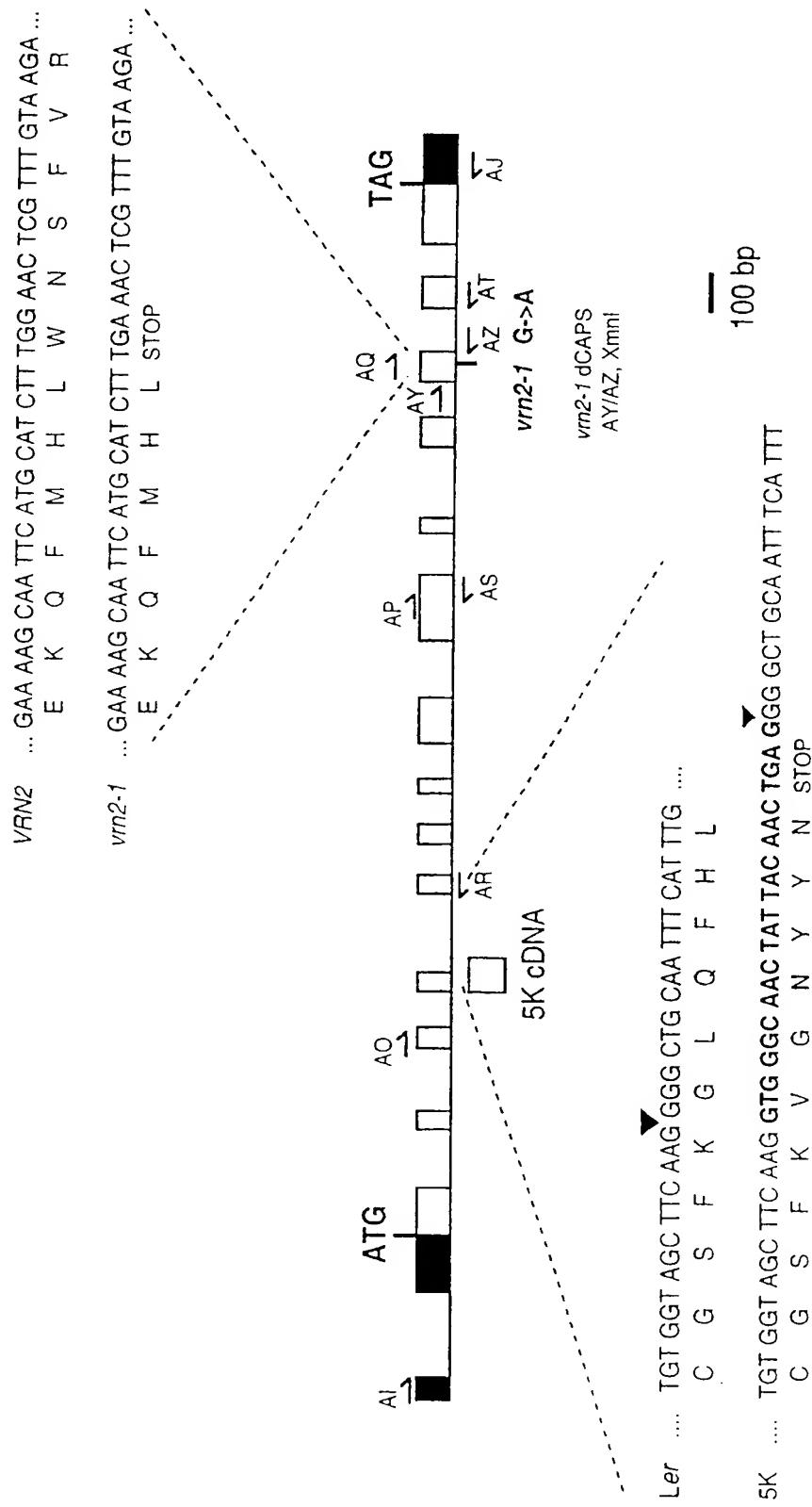


Figure 5



6/12

TTCTTCAATTTCGCTGCTCTCTTACACAGCCAATCGTGTTTCCGAGCTTCAGGCCCTCAATCCAAGACAT	▼	CAAGC	5
TCTATATAAGCATATTGAGAAGAGGGGGCTAATTGTTGACTTATCGCTATGACGTAGGGAAATTCT			80
AATTAGGGAGCCCTCAGAGTTGCACTAACATCAATTGGCTCTTGACGTTGAGTGTAAATTGAAACAAGA			155
			230
ATGTGTAGGCAGAATTGCGCGAAAATCCTCACCGGAGGAAGTGATTCAACTGATGAGAATCTCTTGATATAT			305
M C R Q N C R A K S S P E E V I S T D E N L L I Y			25
TGTAAACCTGTCGACTATATAACATCTTCACCTTCGCTCTAGGCAACCCATCGTTCTCCAAGATGCTTG	▼		380
C K P V R L Y N I F H L R S L G N P S F L P R C L			50
AACTACAAAATTGGAGCAAAGCGAAAAGAAAAGTCAAGATCTACTGGGATGGTAGTTTCAACTATAAGGATIGT	▼		455
N Y K I G A K R K R K S R S T G M V V F N Y K D C			75
AATAACACATTACAGAAAATGGAAGTTAGGGAGGATTGTTCTGTCCATTGGCTCTATGCTATGGTAGCTTC			530
N N T L Q K T E V R E D C S C P F C S M L C G S F			100
AAGGGGCTGCAATTTCATPTGAATTTCATCTCATGATTATTGAATTGAGTTCAAGCTTTCGAAGAATACCAAG	▼		605
<u>K G L Q F H L N S S H D L F E F E F K L F E E Y Q</u>			125
ACAGTTAATGTTCTGTAAAACTTAATTCTTCATATTGAGGAAGAAGGAAGTGATGACGATAAATTGAGCCC	▼		680
T V N V S V K L N S F I F E E E G S D D D D K F E P			150
TTCTCTCTGCTCGAAACCTCGTAAGCGGAGACAAAGAGGTGGCAGAAATAACACCAGGAGACTTAAAGTATGC	▼		755
F S L C S K P R K R R Q R G G R N N T R R L K V C			175
TTTTTACCGTMGGATTCAACCCAGTTAACTAATGGCACAGAAATGGAATCACCTACTTAATGATGGAAACCGT	▼		830
F L P L D S P S L T N G T E N G I T L L N D G N R			200
GGTTTACGGATATCCCCACCCAACAGAGCTTCGGACAATTGAGATGACCAACATTCCACAGCCATAGCC			905
G L G Y P E A T E L A G Q F E M T S N I P P A I A			225
CACTCTCTCTGGACCTGGTGCTAAAGTTATTTGACAAGCGAAGCTGGTCCCTGCTACTAAGACAAGAAG			980
H S S L D A G A K V I L T S E A V V P A T K T R K			250
TTATCTGCTGAGCAGTCAGAGGCTAGAAGCCACCTACTCTTCAGAAACGCCATTCTATCATCTCACAGAGTC	▼		1055
L S A E R S E A R S H L L L Q K R Q F Y H S H R V			275
CAGCCAATGGCCCTTGACCAAGTAATGCTGACCGGGATAGCGAGGATGAAGTCGATGACGATGGCAGATT			1130
Q P M A L <u>E O V M S D R D S E D E V D D D V A D F</u>			300
GAAGATGCCAGATGCTGATGACTTTGCGATGTGAATAAGATGAAAAGCAATTTCATGCATCTTGGAACTCG	▼	▲ vrn-1	1205
E D R Q M L D D F V D V N K D E K Q F M H L W N S			325
TTTGTAAGAAAACAAGGGTTATAGCAGATGGCATATCTTGGGATGTGAAGCATTTCAAGATTTACAGAG	▼		1280
F V R K Q R V I A D G H I S W A C E A F S R F Y E			350
AAAGAGTTCGACCGTTACTCATCACTCTCTGGTGGAGATTGTTTGTAAACTATGGAACCATGGACTT	▼		1355
K E L H R Y S S L F W C W R L F L I K L W N H G L			375
GTCGACTCAGCCACCATCAACAACGCAATACCACCTCGAGAATTGCGTAATAGCTCAGACACCACCAAC			1430
V D S A T I N N C N T I L E N C R N S S D T T T T			400
AACAAACAACACAGTGGATGTCCCAGTGACTCAAACACCAACAACAATAACATTGCGATCATCCAAATGAC			1505
N N N N S V D R P S D S N T N N N N I V D H P N D			425
ATAAAACAACAAGAACATGTTGACAACAAGGACAATAACACCGAGAGACAAAGTAATTAAATAGGAAAATCTCCGG			1580
I N N K N N V D N K D N N S R D K V I K			445
CTTTTATGATACCGATTATCGGATTGTAACCTATTCTCTTAAAAAAATTGTTAGGAGCAAACAAATT			1655
TTATATGTTAGTGTATCAACTGATTACATTAGTAAAAAAATGGATTCTGCTTATAACT			1722

Figure 6
SUBSTITUTE SHEET (RULE 26)



7/12

Figure 7

fca-1 W323 GAAAAGCAATTCATGCATCTT TGG AACTCGTTTGTAAGAA
vrn2.1 TGA STOP

GAAAAGCAATTCATGCATCTTGAACACTCGTTGTAAGAA
CTTNNNNAAG
XmnI site

Diagnostic Primer: VRN2-AZ Antisense
(contains a A and G mismatches at positions 5,7)

Upstream Primer: VRN2-AY 5' TGGCTTCATTAAAGTAGGCAACAGAAAAATGG 3'

Product:

PCR Products

170 bp PCR product for both *fca-1* and *vrm2-1*

GAAAAGCAATTCATGGATCTTGAACTCTCTGTAAGAA
GAAAAGCAATTCATGGATCTTGAAACTCTTCTGTAAGAA

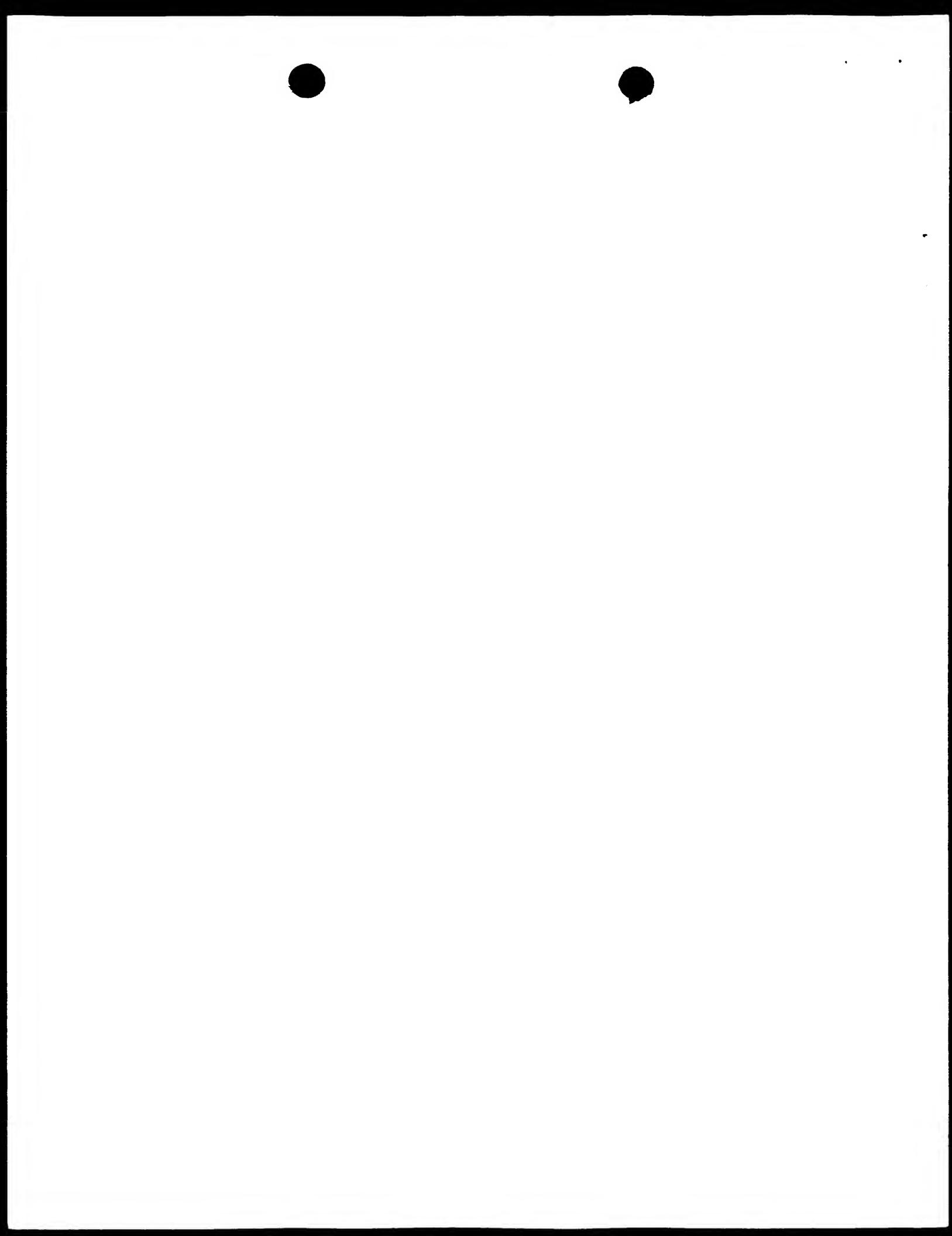
$Xmn1$ digest => *fca-1* no $Xmn1$ site 170 bp
vrn2-1 single $Xmn1$ site 137 bp, 33 bp fragments

SUBSTITUTE SHEET (RULE 26)



8/12

Figure 8a



9/12

Figure 8a continued

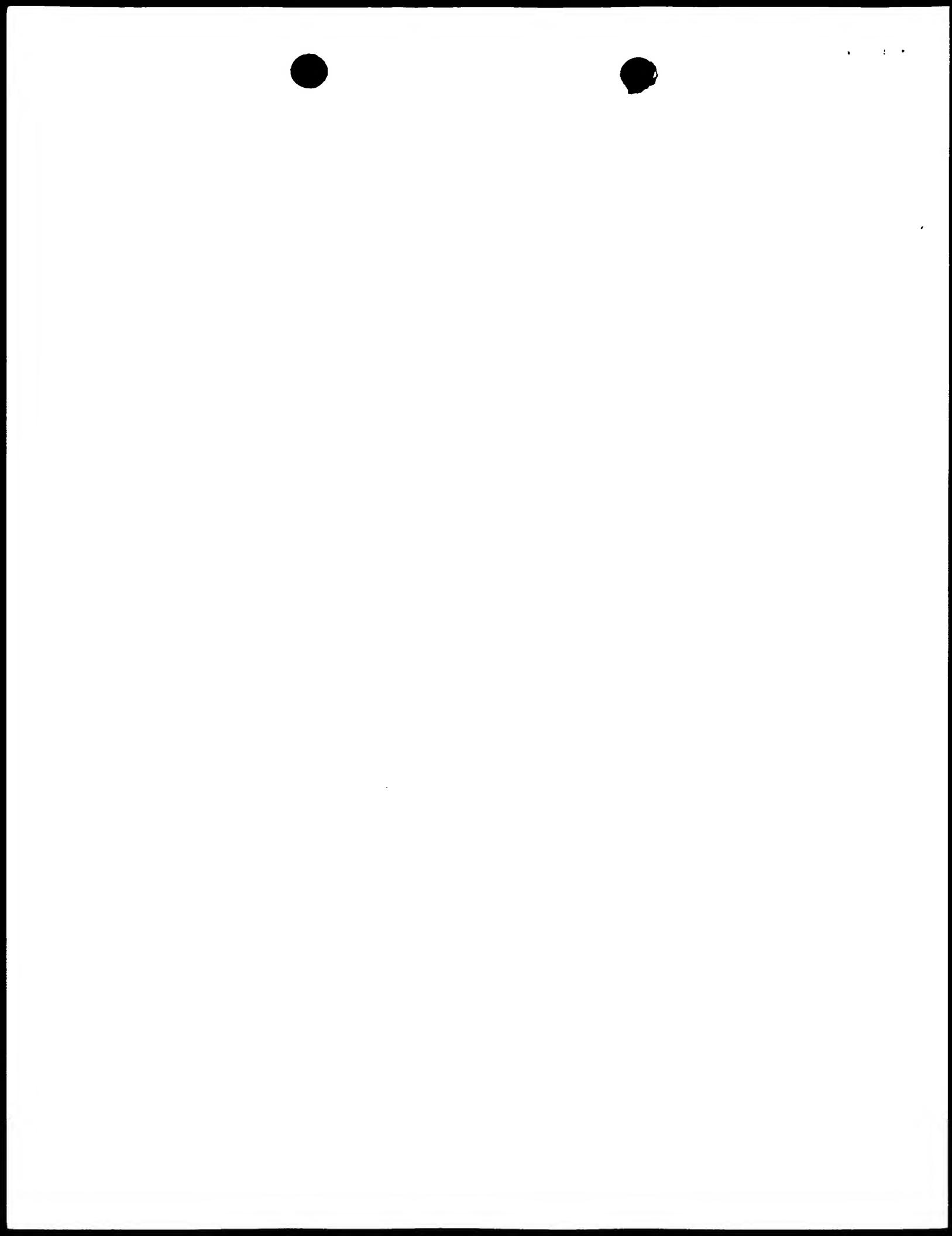
VRN2 Ler	F	E	E	Y	Q	T	V	N	V	S	V	K	L	N	S	F	I	F	E	E	140
AI163743 Prot	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Rice C72616	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
At Hyp 2245035	-	-	-	-	-	-	-	-	D	V	S	I	N	-	-	-	-	-	-	-	557
KIA00160	H	P	K	G	A	R	I	D	V	S	I	N	-	-	-	-	-	-	-	-	557
VRN2 Ler	E	G	S	D	D	K	F	E	P	F	S	L	C	S	K	P	R	K	R	160	
AI163743 Prot	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Rice C72616	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
At Hyp 2245035	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	F	K	K	10	
KIA00160	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	557
VRN2 Ler	R	Q	R	G	G	R	N	N	T	R	R	L	K	V	C	F	L	P	L	D	180
AI163743 Prot	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Rice C72616	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
At Hyp 2245035	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
KIA00160	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	557
VRN2 Ler	S	P	S	L	T	N	G	T	E	N	G	I	T	L	L	N	D	G	N	R	200
AI163743 Prot	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Rice C72616	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
At Hyp 2245035	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
KIA00160	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	557
VRN2 Ler	G	L	G	Y	P	E	A	T	E	L	A	G	Q	F	E	M	T	S	N	I	220
AI163743 Prot	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Rice C72616	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
At Hyp 2245035	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
KIA00160	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	580
VRN2 Ler	P	P	A	I	A	H	S	S	L	D	A	G	A	K	V	I	L	T	S	E	240
AI163743 Prot	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Rice C72616	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	21
At Hyp 2245035	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	74
KIA00160	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	9
VRN2 Ler	N	G	P	V	K	R	T	P	I	T	H	M	D	P	I	C	R	T	E	R	600
AI163743 Prot	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Rice C72616	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
At Hyp 2245035	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
KIA00160	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1



10/12

Figure 8a continued

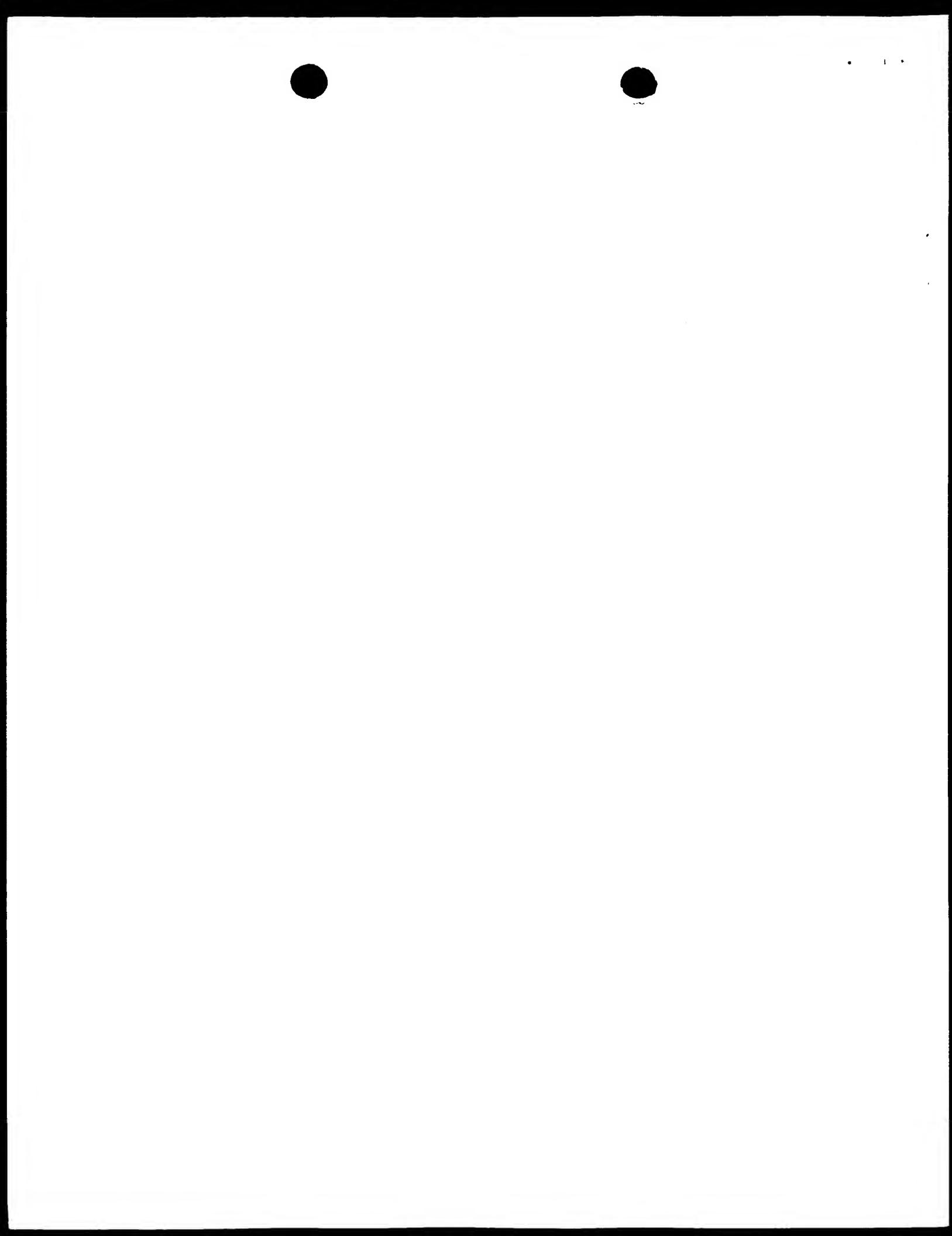
VRN2 Ler	AI163743	Prot	A	V	V	P	A	T	K	T	R	K	L	S	A	E	R	S	E	A	R	S	260
Rice C72616	AI	W	L	Q	F	A	D	M	N	R	N	R	N	S	D	P	M	N	R	N	R	N	40
At Hyp 2245035	A	V	L	Q	F	A	D	M	N	R	N	R	N	S	D	P	M	N	R	N	R	N	93
KIA00160	A	K	V	P	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	21
VRN2 Ler	AI163743	Prot	A	V	V	P	A	T	K	T	R	K	L	S	A	E	R	S	E	A	R	S	620
Rice C72616	A	W	L	Q	F	A	D	M	N	R	N	R	N	S	D	P	M	N	R	N	R	N	279
At Hyp 2245035	A	K	V	P	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	59
KIA00160	A	K	S	M	S	E	F	L	E	S	D	G	E	V	E	C	P	M	A	112	112	112	
VRN2 Ler	AI163743	Prot	H	L	Q	K	R	C	F	Y	H	S	H	R	V	C	P	M	A	112	112	112	112
Rice C72616	A	R	T	-	L	L	H	Q	R	F	H	S	H	R	A	C	P	M	A	112	112	112	112
At Hyp 2245035	H	Y	S	-	P	L	H	K	R	F	H	S	H	R	A	C	P	L	S	41	41	41	41
KIA00160	A	K	S	M	S	E	G	H	N	R	L	Y	F	H	D	T	C	V	L	R	R	R	639
VRN2 Ler	AI163743	Prot	L	E	Q	V	M	S	D	R	D	S	E	D	V	D	-	D	D	V	D	V	297
Rice C72616	A	L	G	A	V	F	M	S	D	R	D	S	E	D	V	D	-	D	D	V	D	V	297
At Hyp 2245035	L	E	Q	V	M	-	D	S	D	R	D	S	E	D	V	D	-	D	D	V	D	V	297
KIA00160	P	Q	E	M	-	E	V	D	S	E	D	E	K	D	-	D	N	D	D	V	A	W	130
VRN2 Ler	AI163743	Prot	A	D	F	E	D	R	Q	M	L	D	D	F	V	D	V	N	K	D	-	E	316
Rice C72616	A	A	D	F	E	D	R	R	M	L	D	D	F	V	D	V	T	K	D	-	E	316	
At Hyp 2245035	A	A	H	L	E	E	S	Q	M	L	N	G	S	M	D	E	N	E	E	D	-	E	316
KIA00160	L	R	E	K	T	I	T	Q	Q	I	E	E	F	S	D	Y	N	E	E	G	-	E	316
VRN2 Ler	AI163743	Prot	K	Q	F	M	H	L	W	N	S	F	V	R	K	Q	R	V	I	A	D	G	336
Rice C72616	K	Q	M	M	M	M	H	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	108
At Hyp 2245035	-	E	R	F	I	K	L	W	N	S	F	V	K	Q	Q	R	I	V	A	D	A	N	154
KIA00160	K	E	V	M	K	K	W	N	L	H	V	M	K	H	G	F	I	A	D	N	N	N	101
VRN2 Ler	AI163743	Prot	H	I	S	W	A	C	E	A	F	S	R	F	Y	E	K	E	L	H	R	Y	356
Rice C72616	A	M	N	H	A	C	M	L	F	V	E	N	Y	G	Q	K	T	R	S	N	K	K	108
At Hyp 2245035	Q	M	N	H	A	C	M	L	F	V	E	N	Y	G	Q	K	T	R	S	N	K	K	154
KIA00160	K	I	P	W	A	C	E	A	F	S	R	L	H	L	Q	E	L	R	S	N	K	K	121



11/12

Figure 8a continued

VRN2 Ler	S S L F W C W R L F L I K L W N H G L V	376
AI163743 Prot		108
Rice C72616		154
At Hyp 2245035	L S L D L C W R Q F M I K Q W D Y G L L	141
KIA00160	- N L C R N F M L H L V S M H D F N L L	733
VRN2 Ler	D S A T I N N C N T I L E N C R N S S D	396
AI163743 Prot		108
Rice C72616		154
At Hyp 2245035	D R V T M N K C N T K I Y H N I S T T N	161
KIA00160	S I M S I D K A V T K L R E M Q K L E	753
VRN2 Ler	T T T T N N N N S V D R P S D S N T N N	416
AI163743 Prot		108
Rice C72616	D D I N N N N T R T T D N M D V V D D	154
At Hyp 2245035	K G E S A S P A N E I T E Q N G T A	181
KIA00160		773
VRN2 Ler	N N I V D H P N D I N N K N N V D N K D	436
AI163743 Prot		108
Rice C72616	I N R D K F S E I N S K E K A L E T D S V S G	154
At Hyp 2245035	N N G F S E I N S K E K A L E T D S V S G	186
KIA00160		793
VRN2 Ler	N N S R D K V I K	445
AI163743 Prot		108
Rice C72616		154
At Hyp 2245035	V S K Q S K K Q K L	186
KIA00160		803



12/12

111
57
85
69
59
65
29
244
82
832
100
269
610
366
590
2734
535
535
8891
420
405
271

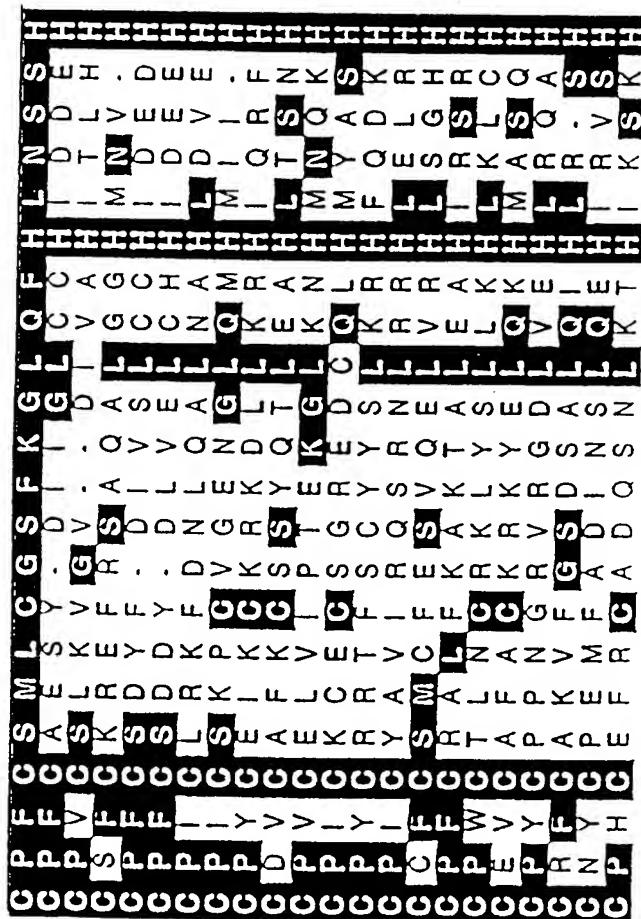


Figure 8b



INTERNATIONAL SEARCH REPORT

Int'l. Application No
PCT/GB 00/00248

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/82 C07K14/415 C12Q1/68 C12N1/20 A01H5/00
C07K16/16

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WILSON A ET AL: "ANALYSIS OF THE MOLECULAR BASIS OF VERNALIZATION IN ARABIDOPSIS THALIANA" SEMINARS IN CELL AND DEVELOPMENTAL BIOLOGY, GB, ACADEMIC PRESS, vol. 7, no. 3, 1 January 1996 (1996-01-01), pages 435-440, XP000609514 ISSN: 1084-9521 page 439 -page 440 --- CHANDLER, J., ET AL.: "arabidopsis mutants showing an altered response to vernalisation" THE PLANT JOURNAL, vol. 10, no. 4, 1996, pages 637-644, XP002139145 the whole document --- -/-	1-10, 15, 17, 46-54, 56, 57
X		1-10, 15, 17, 46-54, 56, 57

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

& document member of the same patent family

Date of the actual completion of the international search

31 May 2000

Date of mailing of the international search report

20/06/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.
Fax: (+31-70) 340-3016

Authorized officer

Holtorf, S

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 00/00248

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LEVY YARON Y ET AL: "The transition of flowering" PLANT CELL, US, AMERICAN SOCIETY OF PLANT PHYSIOLOGISTS, ROCKVILLE, MD, vol. 10, no. 12, December 1998 (1998-12), pages 1973-1989-1989, XP002132682 ISSN: 1040-4651 page 1982, left-hand column ----	1-10, 15, 17, 46-54, 56, 57
X	BEVAN, M., ET AL.: "EU Arabidopsis sequencing project - unpublished" EMBL SEQUENCE DATA LIBRARY, 4 July 1997 (1997-07-04), XP002139150 heidelberg, germany accession no. Z97342 ----	11, 13
A	BANCROFT I ET AL: "THE DEVELOPMENT OF SYSTEMS FOR THE ISOLATION OF GENES FROM ARABIDOPSIS THALIANA BY CHROMOSOME WALKING IN YAC LIBRARIES: TOWARDS THE ISOLATION OF THE FLORAL INDUCTION GENE GCA. UP8.62" JOURNAL OF EXPERIMENTAL BOTANY, GB, OXFORD UNIVERSITY PRESS, vol. 42, no. 238, SUPPL, 7 April 1991 (1991-04-07), page 48 XP002018457 ISSN: 0022-0957 ----	
A	WO 96 38560 A (INNES JOHN CENTRE INNOV LTD ;DEAN CAROLINE (GB); MACKNIGHT RICHARD) 5 December 1996 (1996-12-05) ----	
A	ROBSON PAUL R H ET AL: "Genetic engineering of harvest index in tobacco through overexpression of a phytochrome gene." NATURE BIOTECHNOLOGY 1996, vol. 14, no. 8, 1996, pages 995-998, XP002075914 ISSN: 1087-0156 the whole document ---	
P, X	EP 0 967 278 A (MITSUI CHEMICALS INC) 29 December 1999 (1999-12-29) the whole document ---	11, 12
T	PARKER, J.E., ET AL.: "the Arabidopsis downy mildew resistance gene RPP5 shares similarity to the toll and interleukin -1 receptors with N and L6" EMBL SEQUENCE DATA LIBRARY, 18 November 1999 (1999-11-18), XP002139151 heidelberg, germany accession no. AF180942 ----	

-/--

INTERNATIONAL SEARCH REPORT

Int'l Application No
PCT/GB 00/00248

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
T	<p>SHELDON C ET AL: "The F LF MADS box gene: a repressor of flowering in <i>Arabidopsis</i> regulated by vernalization and methylation"</p> <p>PLANT CELL, US, AMERICAN SOCIETY OF PLANT PHYSIOLOGISTS, ROCKVILLE, MD, vol. 11, no. 3, March 1999 (1999-03), pages 445-458-458, XP002119919 ISSN: 1040-4651</p> <p>-----</p>	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/GB 00/00248

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9638560	A 05-12-1996	AU 709423	B	26-08-1999
		AU 5906096	A	18-12-1996
		CA 2221092	A	05-12-1996
		EP 0832234	A	01-04-1998
		JP 11506001	T	02-06-1999
EP 0967278	A 29-12-1999	AU 3681999	A	13-01-2000